DOCUMENT RESUME

ED 093 248 HE 005 702

AUTHOR	de Wolf, Virginia A.
TITLE	Student Ratings of Instruction in Postsecondary Institutions: A Comprehensive Annotated Bibliography of Research Reported since 1968. Volume 1. Educational Assessment Center Project 264.
INSTITUTION	Washington Univ., Seattle. Bureau of Testing.
PUB DATE	Jun 74
NOTE	86p.

EDRS PRICE MF-$0.75 HC-$4.20 PLUS POSTAGE
DESCRIPTORS *Bibliographies; Educational Research; Evaluation; Faculty Evaluation; *Higher Education; Instruction; *Student Attitudes; Student Opinion; *Student Teacher Relationship; *Teacher Evaluation

ABSTRACT

This 220-item bibliography is the first in a series of reports summarizing research in the evaluation of instruction by students in postsecondary institutions. Articles published since January 1968 and collected by this author before May 1974 furnish the contents. Only research on student ratings of instruction in a postsecondary setting and correlates of such student ratings are included--no pure theory or discussions of possible models for the student ratings of instruction nor research on just administrator's or faculty's ratings of instruction are included. And only research resulting from actual instructional situations are deemed acceptable--results from experimental research settings are not. No quality judgments are made on any of these research findings. The various appendixes with their specific topic headings include: research related to the development, construction and validation of the student rating forms; institutions rated in the articles; specific rating forms employed in the research; student ratings of institution correlated with faculty characteristics and attributes; student ratings of instruction related to similarities between professor/course and student; ideal ratings of instruction and relationships between real and ideal; documentation of changes provoked by the use of student ratings of instruction; and student ratings of instruction correlated with certain characteristics of the course itself. (Author/PG)

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Educational Assessment Center

University of Washington

June 1974

Student Ratings of Instruction in Postsecondary Institutions: A Comprehensive Annotated Bibliography of Research Reported since 1968. Volume I.

Virginia A. de Wolf

Abstract

This bibliography is the first in a series of reports summarizing research in the evaluation of instruction by students in postsecondary institutions. Articles published since January 1968 and collected before May 1974 furnish the contents. 220 abstracts are included. An efficient method for categorizing the contents of all such articles is introduced and its usage in this report is thoroughly explained.

Educational Assessment Center Project: 264

HE 005 702

Student Ratings of Instruction in Postsecondary Institutions: A Comprehensive Annotated Bibliography of Research Reported Since 1968. Volume I.

Virginia A. de Wolf
University of Washington

This bibliography is the first in a series of reports summarizing research in the evaluation of instruction by students in postsecondary institutions. Articles published since January 1968 and collected by this author before May 1974 furnish the contents. This collection of articles will be expanded and updated periodically by, first of all, incorporating all of the past year's research published on the topic and, secondly, by filling in on relevant articles previously overlooked. Hopefully, an additional supplement of the "oldies but goodies" in the field will also be compiled.

Before proceeding with an explanation of the arrangement of this report and the annotations contained within it, a word is needed on what is actually included. Only research on student ratings of instruction in a postsecondary setting and correlates of such student ratings are included--no pure theory or discussions of possible models for the student ratings of instruction nor research on just administrator's or faculty's ratings of instruction are included. And only research resulting from actual instructional situations are deemed acceptable--results from experimental research settings are not. No quality judgments are made on any of these research findings. Caveat emptor. Finally, all works were read in their original forms except for dissertations which are included in this bibliography based upon the value of their abstracted form in the Dissertation Abstracts International (DAS) publication. Of course, the quality of my abstract of a dissertation is directly dependent upon the quality of the summary found in the DAS. The microfilm order number follows the reference citation for each DAS entry as a convenience to those who are interested in a closer inspection of the whole dissertation.

The major portion of the 220 articles included in this volume are from journals (66%). Included under this heading are the dissertations found in the DAS. About eleven percent are papers read at meetings of the Association for Institutional Research, the American Educational Research Association, the American Psychological Association, regional Psychological Associations, and the National Council of Measurement in Education. Institutional research reports and duplicated reports comprise 10% and 8% respectively. The remaining five percent is from books. The ERIC microfiche identification number is included after the reference citation whenever it is available.

The abstracts, arranged in alphabetical order, immediately follow this introductory section. Whenever it existed, the author's (s') abstract is included. In all other instances I have attempted to cover the highlights, as I viewed them, of the research presented. At the end of each abstract is the abstractor's (s') name(s). Finally, there is a series of alphanumeric characters at the end of each entry. These indices refer to items listed in the appendices of this report, and supplement the major points included in the abstract itself by identifying every variable covered in each article.

The various appendices with their specific topic headings are:

A. Research related to the development, construction, and validation of the student ratings forms themselves;
B. Institutions rated in the articles;
C. Specific rating forms employed in the research;
D. Student ratings of instruction correlated with faculty characteristics and attributes;
E. Student ratings of instruction related to similarities between professor/course and student;
F. Student ratings of instruction correlated with student characteristics and attributes;
G. Student ratings of instruction correlated with certain characteristics of the course itself;
H. Comparisons of student ratings of instruction and other types of evaluations of instruction;
I. Ideal ratings of instruction and relationships between real and ideal ratings;
J. Documentation of changes provoked by use of student ratings of instruction; and
K Miscellaneous topics concerning student ratings of instruction.

In the name of thoroughness and accuracy, every important variable was included in the appendices. An individual can make use of these in researching a certain correlate of student ratings of instruction by noting the articles which are indexed to a certain entry and then examining the original reports to see the conclusions of each article. Frequently an author would not include the non-significant correlates in his/her conclusion or abstract of the research. Therefore, I feel that these alpha-numeric characters are a necessary supplement to each entry.

The overwhelming number of items in certain of the appendices, for example Appendices D and F with 38 and 76 entries respectively, might seem a little unwieldy at first glance. The number of items can be whittled down a bit if some of the items are grouped in "interest areas." For instance, a major topic in the correlation of faculty characteristics with student ratings of instruction was the relationship of a faculty member's research productivity with scores on student ratings questionnaires. There are five sub-categories in Appendix D which relate to research productivity:

D6 - research, general research productivity and/or number of publications;

D7 - research, number of awards received (both governmental and institutional);

D11 - research, authorship position (i.e., 1st, 2nd, etc., author on books, articles, etc.);

D31 - research, number of papers read at scientific and professional meetings; and

D32 - research, number of dissertation, thesis or undergraduate honors papers committees chaired.

Student characteristics which have been correlated with student ratings of instruction fall into three major themes. First, grades and grade point averages (GPA) involve nine of the items in Appendix F:

F3 - grade(s), actual course;
F4 - grade(s), type previously received from the same professor;
F5 - grade(s), cumulative gpa;
F23 - grade(s), expected course;
F24 - grade(s), difference between expected and actual;
F25 - grade(s), standardized or common examination;
F36 - grade(s), deserved;
F42 - grade(s), difference between predicted and actual course; and
F68 - grade(s), high school gpa.

Second, eight topics relate to assignments:

F35 - assignments, amount of independent work done;
F43 - assignments - prepared daily or not;
F44 - assignments - prepared for class or not;
F46 - assignments, material not understood - sought explanations or not;

F47 - assignments, minor points not understood - looked up or not;
F49 - assignments, material - made relationships between old and new or not;
F55 - assignments - number of pages of reading read; and
F56 - assignments - number of hours of study per week.

Lastly, variables relating to the students' personal perceptions of the course itself concerns 17 of the entries in Appendix F. Many of these are the following pre-course measures.

F10 - course - how well it fitted into schedule;
F14 - course, perception of classroom atmosphere;
F17 - "course - recommended by another student or not?";
F19 - course, perception of level of learning;
F26 - course, expected difficulty of;
F27 - course, amount of interest in;
F29 - course, preconception of enjoyment to be received from;
F32 - course, expected contribution to planned vocation;
F33 - course, expected contribution to general education;
F34 - course, expected ease of;
F37 - course, effort expended during;
F38 - course, importance of;
F45 - course, participated in voluntarily or not;
F48 - course, class discussions - monopolized or not;
F50 - course, relationships made between this and other courses or not;
F57 - course, reason for selection of; and
F73 - "course - fulfills student's needs?"

The final point I would like to make about the appendices concerns overlapping of certain entries. Because of fine-line distinctions made by certain authors, it is necessary to separate certain highly related topics. Nine such sets are readily apparent: A4 and A15 are concerned with the weighting of student ratings questionnaires, the first on a departmental basis and the second on weighting individual items on a university-wide basis; A5 and J1 deal with the usage of the questionnaires, A5 concerns the calibre of scores when they are used for different purposes while J1 studies the effect of publishing the results upon subsequent ratings of instruction; A12 and F22 on inter-university comparisons--in the first entry the emphasis is on the questionnaires themselves while in the second on the difference in student groups; D10 and G3 concern the field of study, D10 by the professor and G3 as a course characteristic--there is certainly very little difference between the characteristics of instructor and those of the course when it comes to this variable; D16, F41, and G7 all involve different perspectives on teaching methodology--professor, student and course respectively; D19 and D20 both concern teaching practices, the first is general and the second is specific; F11 and F12 the personal values and personality of students, respectively, seem highly related; F31 and F53 involve aspects of the students' student teaching capacity; and F32, F63, and F70 concern the student characteristic of vocational interest.

Adams, H. L. Favorable student evaluations as a function of instructor's age. Improving College and University Teaching, 1973, 21, 72.

During the winter of 1969 the Associated Student Body at Western Washington State College undertook the compilation of a "Course Critique." Instructors in all general education and professional education courses were rated in four areas: "course content and structure," "instructor," "long-range value," and "encourages independent thinking." Numerical ratings were arranged according to instructors' decade of life (20s, 30s, 40s, 50+). While the differences were small there was a consistent decrease with each decade on all four areas. The younger instructors received higher ratings. V. de Wolf. (B2, D3)

Albino, J. E., & Liberty, P. G. Some methodological considerations in the use of factor analysis to determine the dimensions underlying student ratings of courses and instructors. Buffalo, N. Y.: State University of New York at Buffalo, no date. (Duplicated report)

Responses to 33 items of the Course-Instructor Survey were analyzed to determine: (1) comparability of 2 factor structures (one based on individual students' responses versus one based on mean class responses); and (2) the effects of 2 stop-factoring criteria on the factor structures obtained with each of the above data bases (extraction of all factors with latent roots above 1.00 versus the continuous extraction of factors until about 70-75% of the variance is accounted for, or until the reliable variance has been extracted). When the Eigenroot of 1.00 was used as the criterion for stopping factoring, the 6 factors obtained for the data bases were similar. These 6 factors accounted for only 55% of the variance of individual responses, in contrast to 73% of the variance of mean class responses. When factoring was extended, 12 factors were required to account for 70.35% of the variance of individual responses, while 12 factors obtained for the mean class data accounted for 85% of the variance. It was concluded that individual students' responses and mean class responses seemed to provide comparable factor structures in terms of the large, general factors. However, when more specific and meaningful factors were sought, it was recommended that factoring be extended beyond the Eigenroot of 1.00, extracting factors until the variance accounted for was equal to the percentage of reliable test variance, or until the percentage of variance accounted for was increased as much as possible without producing factor fission. V. de Wolf. (A6, C4)

Aleamoni, L. M. Illinois Course Evaluation Questionnaire (CEQ) results interpretation manual Form 66 and Form 32. Research Report #331. Urbana, Illinois: Measurement and Research Division, Office of Instructional Resources, University of Illinois, 1972. (a)

The Illinois Course Evaluation Questionnaire (CEQ) is an instrument used to collect student attitudes towards a course. Its purpose is to enable faculty members to collect evaluative information about their teaching. The data are collected and processed by section but may also be processed by course, department, college, etc. The student responses are anonymous, and two copies of the questionnaire results are returned to the instructor only. He may submit one copy, if he chooses to do so, to his department chairman for consideration in the evaluation of his teaching.

Each instructor's results are compared to an appropriate norm group. Comparisons are made with other instructors of his own academic rank, with those at the same course level (i.e., 100 level = freshmen, 200 level = sophomore, 300 level = junior and senior, 400 level = graduate), with other instructors in his particular department or college, as well as with all courses at that university to all courses that have used the CEQ throughout the United States. Conferences relative to the interpretation of results may be arranged at the instructor's convenience. The evaluation service is free of charge to instructors at the University of Illinois and is available upon request from the Measurement and Research Division, Office of Instructional Resources, 307 Engineering Hall, Urbana, Illinois 61801, (217) 333-3490. Author. (A4, B3, C5)

Aleamoni, L. M. A review of recent reliability and validity studies on the Illinois Course Evaluation Questionnaire (CEQ). Research Memorandum #127. Urbana, Illinois: Measurement and Research Division, Office of Instructional Resources, University of Illinois, 1972. (b)

A review of the research studies since 1965 which have investigated the validity and reliability of the Illinois Course Evaluation Questionnaire (CEQ) was presented. V. de Wolf. (A11, B3, B10, B11, B12, C5, C6, D4, D6, F2, F3, F6, F7, G1, G2, G4, H3, H4)

Aleamoni, L. M. Evaluation by students to identify general instructional problems. In H. R. Smock (Chair), A plan for the comprehensive evaluation of college teaching. Symposium presented at the American Educational Research Association Convention, New Orleans, Feburary 1973. (a)

The development of the Illinois Course Evaluation Questionnaire (CEQ) was documented. V. de Wolf. (A4, A9, B3, C5)

Aleamoni, L. M. The usefulness of student evaluations on improving college teaching. In A. L. Sockloff (Ed.), _Proceedings: The first invitational conference on faculty effectiveness as evaluated by students._ Philadelphia, Pa.: Measurement and Research Center, Temple University, 1973. Pp. 42-58. (b)

Instructors at 2 different institutions who had used the Illinois Course Evaluation Questionnaire (CEQ) during the fall, 1971, and spring, 1972 terms for their courses were scheduled to talk with the author about his/her results. In order to attempt to answer the question of usefulness of student evaluations in improving college teaching, each instructor who had participated in the individual conferences was subsequently followed-up to see if any significant change had occurred in their student ratings in subsequent terms in the same or continuous course(s). Similar CEQ data for instructors who were not able to participate in the individual conferences were available to use as a control group measure. In examining the norm decile changes that took place for the lowest subscale value discussed in the conference, it appeared that the conferences did have a significant effect especially when compared with the control group norm decile changes. V. de Wolf. (A12, B13, B27,JC5, J3)

Aleamoni, L. M., & Spencer, R. E. The Illinois Course Evaluation Questionnaire: A description of its development and a report of some of its results. *Educational and Psychological Measurement*, 1973, *33*, 669-684.

The method of development of the Illinois Course Evaluation Questionnaire (CEQ) was explored up to its current format containing only 50 items. Factor subscores were defined and various methods used for calculating reliabilities of the CEQ were given. The scoring procedure,composition of the normative population, and system of reporting the CEQ results completed the first part of this report. Six studies which have used the CEQ were reviewed. V. de Wolf. (A6, A9, A11, A12, A13, B3, B10, B11, B12, C5, D4, D6, F2, F3, F6, F7, F11, F20, H3, H4)

Aleamoni, L. M., & Yimer, M. An investigation of the relationship between colleague rating, student rating, research productivity, and academic rank in rating instructional effectiveness. *Journal of Educational Psychology*, 1973, *64*, 274-277.

Colleague and student ratings were gathered on a group of 477 instructors and then compared to the instructors' research productivity and academic rank. Colleague and student ratings were not found to be significantly related to the instructors' research productivity. However, colleague ratings were significantly related to academic rank indicating that the reputation of the instructors could be influencing colleague ratings. Authors. (B3, C5, C6, D1, D4, D6, D28, H3)

Aleamoni, L. M., Yimer, M., & Mahan, J. M. Teacher folklore and sensitivity of a course evaluation questionnaire. *Psychological Reports*, 1972, *31*, 607-614. (ERIC: ED076146)

The purpose of this study was to determine if (a) "folklore" about a teacher contributes to his ratings on a course evaluation questionnaire and (b) changes in students' attitudes during the course of instruction can be determined with a questionnaire on course evaluation. Multivariate analysis of variance and discriminant analysis, ideally suited for this type of research, were employed. The results indicated that there were no significant differences in attitudes toward a course in educational statistics between those who took the course in 1967-1968 and those who took it in 1968-1969. This seems to indicate that students do not build a "folklore" about a course based upon the course presented a year earlier. Authors. (A8, A11, B3, C5, D28, J5, J6)

Apt, M. H., & Fahey, G. L. A measurement of college instructor behavior. Scientia Paedagogica Experimentalis, 1968, 5, 32-39.

The specific hypothesis tested was that college instructor behavior could be identified and measured by the use of a graphic rating scale when the rater responded to common, bipolar adjectives. The behaviors would yield, when subjected to factor analysis, clusters of traits which would identify subgroups of college instructors. The instrument was administered to 4,916 undergraduate students in the classes of 112 liberal arts instructors who were categorized by Humanities, Natural Sciences, and Social Sciences academic divisions. Statistical evidence presented in the report does not support the hypothesis. The fundamental conclusion made was that the Osgood Semantic Differential did not discriminate differences among the measures of the three groups tested, and, therefore, it was not possible to make predictions about distinguishing characteristics of the groups. V. de Wolf. (A6, A8, A9, A11, C24, D10)

Asher, E. J., Jr. Differential weighting of teacher evaluation criteria. Paper presented at the meeting of the Association for Institutional Research, Chicago, May 1969.

This study represented an attempt to explore the differential weighting of teacher evaluation criteria at Western Michigan University. The findings of this study indicated that the weighting of evaluation criteria did differ significantly, and that these differences were dependent upon the academic area or group represented. Therefore, it was suggested that teacher evaluation scales or instruments should be tailored to the needs and peculiarities of the particular academic group upon which it was to be used. V. de Wolf. (A4, A9, B14, D10, F7)

Bausell, R. B., & Magoon, J. Expected grade in a course, grade point average, and student ratings of the course and the instructor. Educational and Psychological Measurement, 1972, 32, 1013-1023. (a)

The relationships between expected grade in a course, grade point average (gpa), and instructor-course evaluations were sought. Over 12,000 completed rating evaluations at the University of Delaware were analyzed via multiple-group discriminant analysis and multivariate analysis of variance. The present study found strong, consistent biases in both instructor and course ratings which could be traced to (a) the grade the student expected to receive, and (b) the discrepancy between the student's expected grade and his/her gpa. The relationship between gpa and ratings alone was negligible, and should not be considered an important source of bias. V. de Wolf. (A8, A11, B15, C43, F5, F21, F23, F24)

Bausell, R. B., & Magoon, J. Instructional methods and college student ratings of courses and instructors. The Journal of Experimental Education, 1972, 40(4), 29-33. (b)

This study attempted to determine whether general college student ratings were biased in favor of one type of instructional method or another, and what the structure of these differences was. Ratings on twenty-nine items were compared across five different methods of instruction, utilizing both univariate analyses of variance and multiple group discriminant analysis. The results indicate that the ratings for method differed in three independent ways, two of which were related to students' perceived effectiveness of the instruction. Authors. (A7, A8, A11, B15, F2, F3, F5, F6, F7, G7)

Bejar, I. I., & Doyle, K. O. Generalizability of factor structures underlying student ratings of instruction. Paper presented at the annual meeting of the American Educational Research Association, Chicago, April 1974.

Studied the generalizability of factor structures of student ratings of instruction based on instructors' individual differences. The subjects were instructors from the Humanities, Social Science and Science & Math division who had had their courses evaluated at least twice using the same evaluation questionnaire. The data from the three divisions were factor analyzed and the resulting factor structures compared using Kaiser's procedure. Only one factor ("defines student responsibilities") was found to be stable within and across divisions. The results were explained in terms of the distinction between within-clas covariation and between-class covariation. Authors. (A6, B22, C42, D10, D28)

Berger, W. G., & Cohen, S. H. An empirical model for the formative generalization of a student instructional rating instrument and its relation to various concepts of validity. Ann Arbor: Highway Safety Research Institute, The University of Michigan, no date. (Duplicated report)

This paper describes an extensive research program undertaken to develop an objective instrument designed to measure student evaluation of courses and instructors. Each phase of the research plan was stated in terms of a validity concept: content, face, internal, and construct. Within each phase, procedures were carried out to satisfy the operations underlying the validity. The item frame was generated through a survey of faculty and students on best and worst examplars of teachers and student behavior in the classroom. The items were then subjected to judgments of their apparent relevance to course evaluation. Next, the instrument was analyzed for its factorial structure and its internal consistency. Finally, each evaluative dimension was related to classroom achievement across instructors on a standardized course examination. The entire procedure was offered as a model for generating instruments dealing with course and instructor evaluation. Authors. (A3, A4, A6, A9, B4, C7, F5, F25)

Bills, S. C. The University Evening School of the University of Tennessee: Faculty opinions and teaching performance. Dissertation Abstracts, 1968, 28(10-A), 3961-3962. (Ann Arbor, Michigan: University Microfilms. Order No. 68-3728.)

One section of this dissertation investigated teaching performance of faculty on the University Evening School at the University of Tennessee.

The author hoped to determine if teaching performance of Evening School teachers as rated by students in the teachers' classes differed among academic ranks, to determine if the teaching performance scored varied by students' classifications, and to determine which items in a teacher check list discriminated between effective and ineffective teaching. Some of the major findings were as follows. Graduate assistants were rated the lowest in teacher performance and differed significantly from the other ranks, i.e., from professors, associate professors, assistant professors, instructors, and part time teachers. Adult special students tended to rate teachers higher than other classifications of students. Finally, 15 of 64 items in the teacher performance checklist failed to discriminate between effective and ineffective teaching. V. de Wolf. (B50, D4, D5, D23, F2)

Bittner, J. R. Student evaluation of instructors' communication effectiveness. College Student Survey, 1968, 2, 38-40.

A content analysis of oral communication factors found within the "subjective comments" of the 1968 Course and Instructor Evaluation Report at Purdue University attempted to uncover those factors of oral communication which college students perceived to affect teaching effectiveness, and to determine at which professorial rank and in what schools at Purdue University students' negative comments most often occurred. The research showed that students reacted most favorably to the use of audio-visual aids and very negatively to poor qualities of the speaking voice such as volume, tone or pitch. The highest percentage (76%) of negative comments were attributed to graduate teaching assistants, while instructors had the least (54%). Finally, the greatest percentage of negative comments (87%) belonged to the School of Home Economics and the lowest to the School of Agriculture (46%). V. de Wolf. (A11, B8, C46, D4, D10, D20)

Black, S. Interactions between teaching and research. Universities Quarterly, 1972, 26, 348-352.

In this short article, the author reviewed articles concerning the effect of teaching upon research and research upon teaching. He concluded that there was a deficiency of research in the area of teaching-research interaction. V. de Wolf. (A2, D6)

Blackburn, R. T. Letter on "Research, teaching, and faculty fate." Science, 1971, 172, 1082.

Blackburn commented on the J. R. Hayes' article, "Research, teaching, and faculty fate," by criticizing Hayes' scales and student evaluation item while praising Hayes' ingenuity in devising variables and for utilizing administrator ratings. He also suggested some articles on related studies pertinent to Hayes' research. V. de Wolf.

Blackburn, R. T., & Clark, M. An assessment of faculty performance: Some correlates between administrator, colleague, student, and self ratings. Ann Arbor: Center for the Study of Higher Education, University of Michigan, 1971. (Duplicated report)

The study examines the uncertainties surrounding the evaluation of faculty work performance and the concerns faculty have regarding assessments made about them. Separate evaluations of teaching effectiveness and overall contribution to the college for 45 full-time (85% response rate) faculty members were collected from administrators, faculty colleagues, students, and from the professors themselves.

The very low correlations between the professor and each of his role sets are analyzed. Related research supports a psychological explanation of self misperceptions. The intercorrelations are also examined for their implications. Suggestions are made for improving the evaluation of faculty performance and for mitigating a professor's uncertainties on vital matters affecting his career. Authors. (D3, D8, H1, H2, H3)

Blai, B., Jr. Faculty effectiveness - a 'pilot' study of student evaluation at Harcum Junior College. Report No. IRR 71-37. Bryn Mawr, Pa.: Office of Research, Harcum Junior College, no date.

In 1970-71 academic year over 40% of the students, randomly selected, in each of the college's six divisions responded to a faculty evaluation check list. The author discussed the results of this attempt at student evaluation of faculty effectiveness. V. de Wolf. (B55, C50)

Borland, D. T. A comparative study of instructor ratings by students admitted to a "Disadvantaged Student" Program. Journal of Negro Education, 1973, 42, 187-190.

Of 21 students enrolled in an elective course, ten had met the usual admissions standards of the university and the remainder hadn't. These students completed three evaluative instruments: a general evaluation of the course and instructor; an adjective scale of instructor characteristics; and an instrument designed to measure the total college environment. The major finding of this study was that specially admitted students, most of whom were black and who had been classified as "culturally different or disadvantaged," actually responded in a similar manner to other new freshmen, who were admitted on factors other than racial, social or economic. These two groups, defined as different in cultural background, reflected similar experience in their interpretation of identical input from the same instructor. V. de Wolf. (F18)

Breiter, J., & Menne, J. Measuring teacher performance. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, April 1974.

Step by step procedures for developing a valid and reliable instrument for measuring teacher performance in the classroom are presented. The results of a study applying these procedures are reported. Twenty-one items from a 92-item pool were found to be valid and, by analysis of variance procedures, were shown to discriminate between teachers. These results are interpreted as an indication of a suitable set of procedures to use in evaluating teachers notwithstanding some limitations and measurement problems yet to be solved. Author. (A6, A9, B32, C48)

Bresler, J. B. Letter on "Teaching effectiveness and government awards." *Science*, 1968, *161*, 1160. (a)

This letter was a response to M. Y. Quereshi's letter (Quereshi, 1968) criticizing Bresler's original research reported earlier in *Science* (Bresler, 1968b). V. de Wolf.

Bresler, J. B. Teaching effectiveness and government awards. *Science*, 1968, *160*, 164-167. (ERIC: ED029554) (b)

This report investigated the relationships between publication, success in obtaining government awards, and teaching effectiveness at Tufts University. A search of the literature showed that virtually all comments in the popular literature and most references in professional journals suggested that publication and receipt of support for research somehow detracted from teaching performance in the classroom. The empirical data of this Tufts study did not support these previous conclusions. The students rated as their best instructors those faculty members who had published articles and who had received or were receiving government support for research. V. de Wolf. (B16, D4, D6, D7, D10, D13, G1)

Bridger, J. A. Teacher evaluation with a computerized student opinionaire. *Improving College and University Teaching*, 1973, *21*, 43-45.

The development of a teaching evaluation instrument at Broome Community College was given. V. de Wolf. (A9, B81)

Bridges, C. M., Jr., Brown, B. B., Greenwood, G. E., & Ware, W. B. Analysis of College Teaching, University of Florida. Gainesville, Florida: Institute for Development of Human Resources, College of Education, University of Florida, no date. (Duplicated report)

This short report described the Analysis of College Teaching instrument and also included a copy of the instrument. The authors stated that the purposes of this instrument were (1) to furnish basic descriptive information about the classroom behavior of college teachers, and (2) to provide a framework for establishing norms for making a series of evaluations of teaching performance. This instrument attempted to differentiate description from evaluation. It was not a teacher evaluation instrument *per se*; instead, it was an instrument designed to be used as a *part* of a teaching analysis and evaluation process. The students and faculty who used this instrument were asked to declare values upon which any subsequent evaluations may be asked. The authors concluded their introduction by explaining the make-up of the 5 norm groups and then touched upon the topic of variation in item weights or norms--but they presented no statistical data on either of these topics. V. de Wolf. (A13, A15, C41)

Bridges, C. M., Ware, W. B., Brown, B. B., & Greenwood, G. Characteristics of best and worst college teachers. *Science Education*, 1971, *55*, 545-553.

The present study surveyed stratified samples of student, faculty, and administrators in an attempt to define characteristics of best and worst college teachers. To test for agreement among the three groups in terms of category ranking of stated importance, a Kendall Coefficient of Concordance, W, was computed. The results indicated an extremely high degree of agreement among students, faculty, and administrators as to the perceived relative importance of the stated categories. These findings would suggest then that in the process of developing evaluation instruments which would possibly have meaning for students, faculty, and administrators, items should be developed to assess enthusiasm, preparation and organization presentation, class discussion, interest in students, evaluation (grading, testing assignments), and personality while giving little attention to research and consulting, openness, classroom management, use of A-V materials, performance standards, humility, punctuality, and inquisitiveness. V. de Wolf. (A10, H2, H3)

Bryan R. C. Student rating of teachers. Improving College and University Teaching, 1968, 16, 200-202.

A survey of larger colleges and universities (1,500 or more students) as to past, present, or future use of student ratings of faculty members was reported. 74% of the 416 institutions surveyed replied. V. de Wolf. (A1)

Burnett, C. W., & Badger, F. W. The learning climate in the liberal arts college: An annotated bibliography. Curriculum Series, Number 2. Charleston, West Virginia: Morris Harvey College, 1970.

In this publication the authors attempted to include representative samples of the literature dealing with the learning climate in the liberal arts college. This bibliography was organized into four parts: The Liberal Arts Approach; Curriculum; Teaching Methods and New Media; and The Teaching-Learning Process. The emphasis was on journal literature from 1950 to June of 1969, although some references date back to the 1940's. The last section contained early reference material on student ratings of instructional effectiveness. V. de Wolf. (A2)

Butzhamm, J., & Pfeiffer, M. G. Convergent and discriminant validity of college-level teaching dimensions. Technical Reports Number 9. Philadelphia, Pennsylvania: Psychological Laboratory, La Salle College, 1973.

Six multidimensional scaling analyses based on similarity estimates among 30 college-level teaching tasks, independently made by American, German and Czech psychology students and teachers, resulted in eight teaching dimensions. With the exception of the Czech sample, the status of the rater (whether student or teacher) had little influence on the perceived structure of teaching. Cross-cultural comparisons of the teaching job structures resulted in the following job dimensions common to the perceptions of teachers and students in all of the cultures studied: Control of student behavior, Knowledge dissemination, Information dissemination, Lecturing dynamism, Teacher-student feedback, Counseling contact, Classroom administration, and Use and administration of teaching aids/resources. Convergent and discriminant validity of the eight-factor job structure was demonstrated in a multirater-multitrait matrix. Based on the established construct validity, it was concluded that the eight teaching dimensions may now be employed as subcriteria for the assessment of psychology teaching performance. Authors. (A6, A16, H3) (It should be noted that this report is written in German with abstracts both in German and English.)

Caffrey, B. Lack of bias in student evaluation of teachers. Paper presented at the meeting of the American Psychological Association, Washington, D. C., September 1969.

Student evaluation of teacher performance has been questioned as subject to bias by sex, course grades, over-all grade point ratio (GPR), and personal qualities of the teacher. One hundred thirty-one college students completed a form designed to measure dimensions of teacher performance. A factor analysis of the form showed that student evaluations of desirable teacher characteristics were free from bias. Five stable factors resulted from the analysis: Teaching ability; Feedback to students; Negative attitudes; Student overload; and Structure. Personal qualities of the teacher, sex of the student, grades in the class, and overall GPR did not influence student assessments of teacher performance. Author. (A6, A7, C8, F3, F5, F6)

Caffrey, B. Use of multiple regression analysis and factor analysis to design a brief, objective assessment of instructor performance. Paper read at annual meeting of the Southeastern Psychological Association, New Orleans, April 1973.

A 50-item Student Report Form to evaluate an instructor on various dimensions of teaching performance was reduced to a brief 18-item scale by a series of factor analyses. The final factor analysis assessed only the 3 most consistently derived factors: I. Proficiency or Skill; II. Difficulty of the Course; and III. Structure or Organization. Item scores were combined according to regression weights derived from optimal predictors of standardized factor scores obtained from factor analyses. Factor scores were standardized with a mean of 75.0 and a standard deviation of 15.0. Thus a teacher who received a standardized factor score above 90.0 might be compared with a student who received a grade of 90.0 or better. This technique of scoring, while statistically rigid, can be easily understood by the college teacher who is not familiar with statistics. V. de Wolf. (A6, A9, C31, D12, F3)

Canter, F. M., & Meisels, M. Cognitive dissonance and course evaluation. Improving College and University Teaching, 1971, 19, 111-113.

A group of college students in Introductory Psychology classes were queried concerning their expectations about the difficulty and value of the course at the beginning and again at the end of this experience. According to the cognitive dissonance theory, a positive relationship would be expected between ratings of course difficulty and course value. The results showed that cognitive dissonance appears to be present as a factor in the subjects' expectancies, though it was not the only factor or even the major one. The students did expect that the course difficulty would also be an indication of its value. However, as a result of going through the experience of the course, this expectation was attenuated and the relationship between judged value and difficulty became almost a negligible factor. Concomitantly, the previously de-emphasized factor of obtained academic grade became more important, indicating that the "reward" of a higher grade influenced opinion of the course value. V. de Wolf. (F3, F23, F26, F27)

Carter, R. E. The effect of student characteristics on three student evaluations of university instruction. Dissertation Abstracts International, 1969, 30(2-A), 486. (Ann Arbor, Michigan: University Microfilms. Order No. 69-4735.)

Purposes of the study were: (1) To examine the intercorrelation among item and total scores on the following three student evaluations of instruction instruments: the Indiana University Student Evaluation of Instruction (IU), the Faculty Evaluation Survey (FE), and the Purdue Instructor Indicator (PIPI); (2) To examine the influence of the student's seriousness of academic intent as rated by Resident Assistants, the student's grade point average (GPA), academic discipline, status, and institutional background upon the student's evaluation of a university course as measured by the above named instruments; and (3) To develop a revised instrument for student evaluation of university instruction. Some of the results follow. Total scores of the 3 instruments were significantly intercorrelated ($P > .01$). The student's seriousness of academic intent did not significantly affect ($P < .05$) his evaluation of instruction as measured by the Resident Assistant ranking of the students, but this ranking did significantly predict ($P > .01$) the GPA of the students. GPA had significant effect on the IU but not on the FE or PIPI. The student's academic discipline did not significantly affect the results of the FE but did affect the results of the IU and PIPI. Student status did not significantly affect ($P < .05$) the results of the IU and FE, but freshmen rated the course significantly higher ($P > .01$) than did sophomores on the PIPI. The student's institutional background significantly affected ($P > .01$) his evaluation on all 3 instruments. V. de Wolf. (C1, C2, C44, F2, F5, F7, F54)

Cash, B. A classroom feedback system. Improving College and University Teaching, 1972, 20, 291.

The author described a critique system which allows students to participate anonymously in his communication classes. Author. (A9)

Cassel, R. N. A student course evaluation questionnaire. Improving College and University Teaching, 1971, 19, 204-206.

The development of the University of Wisconsin-Milwaukee Student Course Evaluation Questionnaire (CEQ) and a preliminary report using this CEQ were included in this article. Factor analysis of the 50 item CEQ identified four factors. Intercorrelations of these four factors with one another and with the Reading Validity Index (RVI), an instrument developed to determine internal consistency of rater responses, tended to be somewhat higher than desirable for maximum effectiveness. V. de Wolf. (A5, A9, A11, A13, B17, C9, C10)

Centra, J. A. Two studies on the utility of student ratings for improving teaching. SIR Report No. 2. Princeton, N. J.: Educational Testing Service, 1972.

This report is presented in two parts. In the first part, the effectiveness of student ratings in improving college teaching is investigated. An experimental study was conducted at five diverse colleges to investigate the question: To what extent do college teachers alter their instructional practices after receiving feedback from students? The "treatment" in this study purposely paralleled what is typically done at most colleges that use student ratings--that is, the results were seen only by the individual teacher, and the interpretation of the ratings as well as what to do about them were left up to the teacher.

In the second part of this report, teacher self-ratings are compared to student ratings in order to investigate the extent to which instructors learn something from students about their teaching. Because this second part will eventually be published separately, descriptions of the instruments, sample, etc., are repeated. Author. (A11, A12, C3, D1, D2, D10, F22, H1, J2, J3, J5, K2)*

*Both of these articles have been published by Centra and are included in this bibliography. The first is under the title of "Effectiveness of student feedback in modifying college instruction" and the second as "Self-ratings of college teachers: A comparison with student ratings."

Centra, J. A. Effectiveness of student feedback in modifying college instruction. Journal of Educational Psychology, 1973, 65, 395-401. (a)

An experimental study was conducted at five colleges to investigate the extent to which college teachers modify their instructional practices after receiving student feedback. Variables included teaching experience, sex, and self-ratings of the instructor, as well as course subject area. On the basis of equilibrium theory, a major hypothesis of this study was that student ratings would produce changes in teachers who had rated themselves more favorably than their students had rated them. Results of a regression analysis generally supported this hypothesis. A second conclusion of the study was that additional time (i.e., more than a half of a semester), along with comparative data to help the individual teacher interpret his feedback, also helped produce modest changes in teachers' instructional practices. Author. (C3, D1, D2, D10, F22, H1, J2, J3, J5)

Centra, J. A. The relationship between student and alumni ratings of teachers. Research Bulletin 73-39. Princeton, N. J.: Educational Testing Service, 1973. (b)

Student and alumni ratings for 23 teachers were found to correlate .75 (somewhat less for teachers rated only by graduates of their department). This substantial agreement between current students and alumni (of five years) regarding who have been effective or ineffective teachers suggests that student ratings are fairly permanent and do, at least in part, reflect overall, long-term effects of instruction. Author. (A11, E3, H5)

Centra, J. A. Self-ratings of college teachers: A comparison with student ratings. Journal of Educational Measurement, 1973, 10, 287-295. (ERIC: ED069701) (c)

College teachers' self-ratings were investigated in this study by comparing them to ratings given by students. The sample consisted of 343 teaching faculty from five colleges; these teachers, as well as the students in one of their classes, responded to a 21-item instructional report questionnaire. Teacher self-ratings had only a modest relationship with the ratings given by students (a median correlation of .21 for the items). In addition to the general lack of agreement between self and student evaluations, there was also a tendency for teachers as a group to give themselves better ratings than their students did.

Discrepancies between individual teacher ratings and ratings given by the class were further analyzed for: (a) sex of the teacher (no difference found); (b) number of years of teaching experience (no difference); and (c) subject area of the course (differences noted for natural science courses vs. those in education and applied areas). Author. (C3, D1, D2, D10, F22, H1)

Centra, J. A. The student as godfather? The impact of student ratings on academia. Research memorandum 73-8. Princeton, N. J.: Educational Testing Service, 1973. (d)

In this paper the author has attempted to discuss the effects or possible effects of student evaluations on academia. To draw support for his opinions he has included a short review of validity studies which have been conducted at ETS. V. de Wolf. (D1, D2, D10, F7, H1, H3, J2, J3)

Centra, J. A. The student as godfather? The impact of student ratings on academia. In A. L. Sockloff (Ed.), Proceedings: The first invitational conference on faculty effectiveness as evaluated by students. Philadelphia, Pa.: Measurement and Research Center, Temple University, 1973. Pp. 25-41. (e)

See preceding abstract of this article. V. de Wolf.

Centra, J. A. College teaching: Who should evaluate? Findings, a Quarterly of ETS Research in Postsecondary Education, 1974, 1(1), 5-8.

In this article, the author reviewed the results of 2 of his studies. The purposes of the first study reported were to assess the reliability of colleague evaluations of instructor performance in the classroom and to compare colleague and student ratings of the same instructors. The Student Instructional Report (SIR) was used to gather student reactions to instruction and 16 items from the SIR were selected for the colleague ratings instrument. Each professor being evaluated was observed and rated twice by each of 3 colleagues. The analysis of the colleague ratings showed that there was not a great deal of agreement among colleague raters. In viewing the level of agreement between students and colleagues on the 16 items from the SIR, a correlational analysis was performed. There was reasonable agreement on 6 of the 16 items but there was little convergence between the 2 sets of ratings for the other 10 items. In sum, then, this study yielded strong evidence that, when it came to evaluating certain aspects of college teaching, students were likely to do a better job than colleagues, given the practical limits on faculty observation. The second study which was reported provided evidence that student ratings could help instructors improve their teaching practices, even if the results were seen only by the individual teacher. V. de Wolf. (A7, A11, C3, J3, H1, H3)

Centra, J. A., & Linn, R. L. Student points of view in ratings of college instruction. Research Bulletin 73-60. Princeton, N. J.: Educational Testing Service, 1973.

This study investigated student points of view in their ratings of specific courses and instructors by analyzing student responses within each of three classes separately, and then sought to generalize the results further by additional analyses with 300 students randomly selected from 402 classes in five colleges. Different points of view were found for student ratings of course examinations, textbooks and supplementary readings, and class discussions. These various points of view were moderately related to such student characteristics as grades and sex, although not in all three of the classes studied. This last point underscores the importance of the context (i.e., the particular course) in understanding or interpreting the meaning of student ratings. Authors. (A6, A8, C3, E4, F2, F5, F6, F7, F23)

Choy, C. The relationship of college teacher effectiveness to conceptual systems orientation and perceptual orientation. Dissertation Abstracts International, 1969, 30(5-A), 1860-1861. (Ann Arbor, Michigan: University Microfilms. Order No. 69-19, 212.)

The study was designed to investigate college teacher effectiveness by comparing certain global ratings of faculty members on a number of criterion measures, and their conceptual systems orientation and perceptual orientations. 32 faculty members of the Department of Psychology at Colorado State College completed 2 instruments, the "This I Believe" instrument and the Perceptual Orientation Scale, to measure the theoretical constructs of Conceptual Systems Orientation and Perceptual Orientation. Teacher effectiveness was determined by ratings from 4 sources: self; administrator; colleagues; and students. It was concluded that faculty subjects' characteristic ways of perceiving other people, self, and their task and their Conceptual Systems Orientation were related to their effectiveness as college teachers, but that whether or not faculty subjects were full-time or part-time, or male or female was not related. Self ratings of effectiveness were very different than ratings of effectiveness by administrator, colleagues, and students. These two were regarded as separate dimensions of college teacher effectiveness. Lastly, faculty subjects' Conceptual Systems Orientation was related to their characteristic ways of perceiving others, self, and task. V. de Wolf. (B7, D1, D12, D19, D23, D26, H1, H2, H3)

Clark, M. J., & Blackburn, R. T. Faculty performance under stress. In A. L. Sockloff (Ed.), Proceedings: The first invitational conference on faculty effectiveness as evaluated by students. Philadelphia, Pa.: Measurement and Research Center, Temple University, 1973. Pp. 233-252.

The major focus of this research report was on the stresses faculty members felt as they conducted their work and the relationship between these conflicts or pressures and their performance as classroom teachers. The performance measures used in the study were ratings of teaching effectiveness by faculty colleagues and also ratings of teaching effectiveness by students in each of the professor's classes. The first hypothesis stated that high stress would negatively affect work performance, or rated teaching effectiveness. Though student ratings were somewhat lower for faculty members who reported that they felt a lot of pressure and conflict, the differences between the high and low stress groups were not statistically significant and this hypothesis was rejected. Students and faculty had highly similar patterns of assessment concerning faculty teaching effectiveness. There was general agreement on relatively higher or lower ratings as well as on the effects of stress and the moderation of this effect by personal characteristics. V. de Wolf. (D2, D3, D4, D12, D15, D17, D21, H3)

Cohen, S. H., & Berger, W. G. Dimensions of students' ratings of college instructors underlying subsequent achievement on course examinations. Proceedings, 78th Annual Convention, APA, 1970, 605-606.

The present study examined the correspondence between an instructor's mean rating on various dimensions of teaching effectiveness and subsequent achievement by his/her class on a standardized, objectively scored test. Completed ratings on the Michigan State University Student Instructional Rating Report (SIRR) were available for 25 instructors teaching sections of a basic natural science course. Mean scores on the five subscales of the SIRR and the total scale were correlated with mean pre-course cumulative grade-point average for each section (PreGPA) and with mean test performance. None of the subscales were significantly correlated with PreGPA. Moreover, the correlation between mean PreGPA and mean test achievement was negligible (r = -.06). Significant correlations did emerge between test achievement and, in order of decreasing magnitude, student interest, student-faculty interaction and course organization. V de Wolf. (B4, C7, F5, F25)

Cope, R. G., McMillin J. G., & Richardson, J. M. A study of the relationship between quality instruction as perceived by students and research productivity in academic departments. Seattle, Wash.: College of Education, University of Washington, 1972. (Duplicated report)

This report presents the findings from a study of the relationship between quality teaching and reputation for research and publication in seventeen academic departments at the University of Washington. A specially prepared student rating of teaching form was administered in the spring of 1972. The results from the student ratings of instruction were compared with these departments' national ranking in the most recent American Council on Education rating of graduate programs.

The results did not indicate a relationship exists between the student perception of quality of instruction and the reputation departments have among peers for research and publication. The results did show, however, that the relationship between student ratings and research reputation differed according to field of study, and that the results of current research do enter into the instruction received in advanced courses. Authors. (B1, F2, G3, K1)

Costin, F. A graduate course in the teaching of psychology: Description and evaluation. Journal of Teacher Education, 1968, 19, 425-432.

Student ratings received by two groups of psychology teaching assistants were compared. One group had participated in a seminar on the teaching of psychology and the other group hadn't. The ratings of assistants were obtained at three different times during the academic year: middle of the first semester, end of the first semester, and end of the second semester. Assistants who had participated in the seminar made a significantly greater gain in the first semester on the factor of student-teacher "rapport." The adjusted mean scores of the seminar group were also higher than those of the nonseminar group for the other four factors of "skill," "structure," "feedback," and "interaction," but none of these differences were significant. A comparison of some of these teaching assistants during two consecutive semesters was made even though there was a reduction in the total number of assistants in each group. Comparisons of adjusted mean ratings at the end of the second semester revealed two significant differences: one for "feedback" and the other for group "interaction." In each instance, the seminar group received higher mean ratings than did the nonseminar group. V. de Wolf. (B3, C8, J5, K4)

Costin, F., Greenough, W. T., & Menges, R. J. Student ratings of college teaching: Reliability, validity, and usefulness. Review of Educational Research, 1971, 41, 511-535.

The purpose of this article was to review extensively and critically empirical findings concerned with the reliability, validity, and usefulness of student ratings. It also reported the results of a recent survey carried out by the present authors, in which students at the University of Illinois were asked to express their opinions about the use of student rating forms in assessing classroom instruction. V. de Wolf. (A2, A3, A6, A10, A11, A12, B1, B3, B8, B18, B19, B20, B21, B22, D1, D2, D4, D6, D7, D10, D12, D16, D28, F1, F2, F3, F7, F12, F16, F23, F25, F27, F62, G1, G5, H1, H2, H3, I4, J3, J7, K4)

Crawford, P. L., & Bradshaw, H. L. Perception of characteristics of effective university teachers: A scaling analysis. Educational and Psychological Measurement, 1968, 28, 1079-1085.

From an analysis of 300 student themes describing university teacher characteristics which the students considered essential for "effective teaching," 13 different descriptive statements emerged. These descriptions were randomly arranged for paired comparison scaling and presented to 158 students, 50 university faculty, and 30 administrators and department chairmen/women. Scale values were computed for the 13 descriptive statements for each of 10 groups. Statistically significant coefficients of agreement were found for each matrix which indicate perceptual agreement among the judges within each group. Comparisons of each matrix of judgments with the "population" matrix revealed significant differences between values of descriptive statements assigned by different groups of judges. It was concluded, therefore, that characteristics of effective teachers may be quantified and studied scientifically, but that these characteristics should be interpreted only in light of who is doing the judging. V. de Wolf. (A9, H2, H3, H8)

Crittenden, K. S., & Norr, J. L. Student values and teacher evaluation: A problem in person perception. Sociometry, 1973, 36, 143-151.

Student evaluation of teaching is treated as a special case of person perception in an effort to assess the effect of student values on the evaluation process. The specific model proposed is that a student's overall evaluation of an instructor is an additive combination of evaluations of individual aspects of teaching behavior weighted by the student's estimation of the relative importance of these aspects to good teaching. Two hypotheses derived from the model are supported by teacher evaluation data collected in natural college classroom settings. These results strengthen support for processes of impression formation posited in experimental studies of person perception and highlight the importance of assessing student expectations and values as part of the teacher evaluation process. Authors. (A9, A15, F39)

Crooks, T. J., & Smock, H. R. Student ratings of instructors related to student achievement. Urbana, Illinois: Office of Instructional Resources, University of Illinois, 1974. (Unpublished paper)

In this study, global student ratings of their instructors were correlated with 8 student achievement variables, using class means for 29 sections of an introductory college physics course. Correlations were corrected for initial differences among classes by partialling out selected pretest variables. Of the 48 correlations and partial correlations reported, 47 were significantly greater than zero. These findings, together with those of previous research in the area, indicate a generally positive relationship between global instructor ratings and student achievement. A subsidiary analysis suggests that the magnitude of correlation found in studies of this type may be strongly dependent on the particular sample of instructors included in the study. Authors. (A11, F3, F16, F21)

Dedrick, C. V. L. The relationship between perceptual characteristics and effective teaching at the junior college level. Dissertation Abstracts International, 1973, 33(8-A), 4170-4171. (Ann Arbor, Michigan: University Microfilms. Order No. 73-550.)

See abstract of the following report by Dedrick (Dedrick, no date) for a summary of this dissertation. V. de Wolf.

Dedrick, C. V. L. The relationship between perceptual characteristics and effective teaching at the junior college level. Cedar Falls, Iowa: University of Northern Iowa, no date. (Duplicated report) (ERIC: ED070430)

The major hypothesis of this research was that there would be a significant positive relationship between student ratings of faculty effectiveness and certain aspects of perceptual organization as measured by the Thematic Apperception Test (TAT). The subjects were 32 junior college instructors. Two student ratings questionnaires were administered to students of all 32 instructors: (1) the Purdue Instructor Performance Indicator (PIPI); and (2) the Teacher Rating Scale (TRS). Multiple stepwise regression correlations were computed between the 6 perceptual dimensions inferred from TAT protocols and each of the student ratings forms. Three perceptual dimensions emerged in the same relative order in terms of the relationship between perceptual variables used as predictors and both the PIPI and the TRS: teacher's perceptions of the helping relationship as freeing rather than controling; teacher's perception of others as able rather than unable; and teacher's perception of self as adequate rather than inadequate. V. de Wolf. (C11, C44, D12, D26)

Derry, J. O., Bache, W., Golden, G., Pullen, S., Seibert, W., Starry, A., Sussman, L., Van Horn, J., & Wright, G. The Cafeteria System: A new approach to course and instructor evaluation. Institutional Research Bulletin 74-1. Lafayette, Indiana: Measurement and Research Center, Purdue University, 1974.

Of the purposes for conducting course and instructor evaluation, diagnosis of teaching strengths and weaknesses is considered important to programs of instructional development. Accordingly, the staff of the Measurement and Research Center has designed and recently implemented at Purdue a highly flexible, diagnostic, and computer-assisted evaluation system. The system, called "Cafeteria," permits individual instructors to select items from a catalog and thus to build tailor-made rating instruments. Rating instruments are produced in specified quantities by computer, administered to classes and returned for scoring. Item response characteristics for each catalog entry are retained and norms updated each semester. In addition to selected items, each rating scale contains an abbreviated standardized "core" which is added automatically. This core serves at least two purposes: 1) to provide a baseline common to all faculty for diagnostic interpretation and 2) to retain a secondary capability for supporting administrative decision-making.

Cafeteria is an evaluation system uniquely suited to a wide-range of instructional strategies and teaching situations and thus it is responsive to increased demands for a method to assess the impact of innovative teaching. With recent developments in computer technology, especially the growing availability of mini-computers, Cafeteria represents a viable approach to course and instructor evaluation for educational institutions. Authors. (A4, A9, B8, C47)

Deshpande, A. S., Webb, S. C., & Marks, E. Student perceptions of engineering instructor behaviors and their relationships to the evaluation of instructors and courses. American Educational Research Journal, 1970, 7, 289-305.

674 undergraduate students in a university Mechanical Engineering department rated their teachers the week following midterm exams on the Teacher Description Instrument (TDI). Fourteen first-order factors were extracted from the TDI which represent the dimensions of the domain of teacher behaviors. In order to examine the interrelationship, if any, among the first-order factors, a second-order factor analysis of the 14 factor scores was performed. Four second-order factors resulted. Finally, to examine the relationships between the dimensions of teaching behaviors and the criteria, the first- and second-order factor scores were correlated with the three criterion variables: emphasis placed on theory and practice in the course, the overall value of the course, and instructor's teaching ability. The results were thoroughly discussed. V. de Wolf. (A6, A9, A11, C12)

Domino, G. Interactive effects of achievement orientation and teaching style on academic achievement. Journal of Educational Psychology, 1971, 62, 427-431.

This study tests the hypothesis that there is an interaction between a student's achievement orientation and the teaching style he is exposed to, which differentially affects both the amount of learning that takes place and the degree of expressed satisfaction with the scholastic environment. One hundred students, selected because of their extreme scores on the Achievement-via-Conformance and Achievement-via-Independence scales of the CPI, were assigned to introductory psychology sections taught in either a conforming or an independent manner. An analysis of their scores on a final examination consisting of multiple-choice items and essay questions, as well as their ratings of teacher effectiveness and course evaluation, indicated a clear interaction effect. Students taught in a manner consonant with their achievement orientation obtained significantly higher means on the multiple-choice items, on factual knowledge ratings of their essay answers, and on their ratings of teacher effectiveness and course evaluation, than their peers taught in a dissonant manner. Author. (D16, F3, F5, F12, F16, F25).

Dow, D. Letter on "Research, teaching, and faculty fate." Science, 1971, 172, 1082-1083.

Dow considered that the information presented by Hayes in his article, "Research, teaching, and faculty fate," gave no good information on causality of the teaching-and-research relationship. The key point overlooked, according to Dow, was a clarification of what constituted good teaching. He stated that "studies of this sort might be more useful if they employed more realistic indices of research achievement, such as the number of times an author's paper is cited rather than how many papers he has produced." V. de Wolf.

Doyle, K. O., Jr. Construction and evaluation of scales for rating college instructors. Dissertation Abstracts International, 1972, 33(5- A), 2163. (Ann Arbor, Michigan: University Microfilms. Order No. 72-27,746.)

The author reported the construction, development, and evaluation of an instructor ratings scale at the University of Minnesota. The final 107 items yielded 4 General Scales which were derived from factor analysis. These 4 General Scales were further broken down into 16 Specific Scales. Intercorrelations among the General Scales ranged from .54 to .82, the Hoyt internal consistency coefficient from .90 to .96, and the specific variance from .26 to .64; among Specific Scales, correlations were .36 to .84, internal consistency coefficients .61 to .92, and specific variance .12 to .79. The variance specific to each of the scales seemed sufficient to justify their retention, in spite of the rather high intercorrelations. Scales were, for the most part, transferrable across the student characteristics studied: year in college, grade-point average, sex, expected grade, and curriculum. V. de Wolf. (A6, A9, B22, F2, F5, F6, F7, F23)

Dwyer, F. M. Selected criteria for evaluating teacher effectiveness. Improving College and University Teaching, 1973, 21, 51-52.

From a review of 62 studies on the characteristics and relationships of selected criteria for evaluating teacher effectiveness, the author put forth generalizations derived from the research cited. V. de Wolf. (A2, A11, C43, D1, D2, D3, D6, D9, D14, D18, D19, F2, F6, F16, G1, H5)

Elliot, C. K. Longitudinal use of student-constructed teacher evaluation form. British Journal of Educational Psychology, 1969, 39, 309-313.

A simple teacher evaluation form was developed from variables used by students when considering teachers. Results indicated consensus amongst a large group of students about teacher strengths and weaknesses. Similar profiles emerged over each of three years, but median ratings indicated an overall improvement in performance. Profiles were shown to discriminate between different lecturers. Author. (A9)

Elmore, P. B., & LaPoint, K. A. Sex factors in the evaluation of college instructors. Technical report 3.1-73. Carbondale, Illinois: Counseling and Testing Center, Southern Illinois University, 1973.

This study was conducted to determine if the evaluations of faculty are affected by the sex of the faculty member being evaluated and the sex of the student doing the evaluation. In order to determine the effect of sex factors on evaluation, only the twenty questions on the Instructional Improvement Questionnaire that directly evaluate instructor performance were used. The research question was: Are male and female faculty members evaluated the same by male and female students on each of twenty items measuring instructor performance?

The research question was answered by using the two-factor analysis of variance procedure. The findings showed no interaction between faculty sex and student sex on each of the twenty items. In general, there were no differences on each of the twenty items between (1) mean ratings given male instructors and female instructors and (2) mean ratings given by male students and female students to all instructors. However, male instructors did receive higher ratings on "spoke understandably" than female instructors and female instructors received higher ratings on "promptly returned homework and tests" than male instructors. It was also found that female students rated instructors higher than male students rated them on "specified objectives of the course." Authors. (B23, C13, D1, F6)

Elsmore, T. F. Letter on "Research, teaching, and faculty fate." Science, 1971, 172, 1083.

"It is my opinion the his [Hayes] measures of 'teaching ability' were entirely inadequate," commented Elsmore in reference to Hayes' article, "Research, teaching, and faculty fate." Elsmore contended that Hayes showed that two measures of student evaluation of teaching do not covary but then tried to relate these measures to other variables. V. de Wolf.

Feldhusen, J., & McDaniel, E. Using a model of instruction as the base for course and instructor evaluation. In D. N. Braunstein (Chair), Toward improved models of college teaching evaluation: Instructor, student, and organizational influences. Symposium presented at the American Psychological Association, Honolulu, September 1972.

The purpose of this paper was to review a number of studies on instructor evaluation which have been carried out recently at Purdue University and to propose some new methodologies, based on the use of instructional models, for course and instructor evaluation. V. de Wolf. (A2, B8, C43, D16, E2, F11, F12, H1, H7, H8, I1, I4, K2)

Feldhusen, J. F., & Starks, D. D. Bias in college students' ratings of instructors. College Student Survey, 1970, 4, 6-9

The present research was concerned with this question: Are students' ratings of instruction at the end of a college course determined by their general attitudes towards courses and instructors? 257 students enrolled in a course in educational psychology served as subjects. The Purdue Rating Scale for Instruction (PRSI) was revised slightly both with respect to the directions given and in the wording of the items to cause students to focus on his college courses and instructors in general rather than on a specific course or instructor. This revised rating scale was administered near the beginning of a semester. The regular PRSI was then given at the end of the semester as an assessment of the instructor and the course in educational psychology in which the students were enrolled. The correlations between students' general attitudes toward instruction and toward the specific course they were just completing were, even though significant, quite low. Therefore, it seemed reasonable to assume that students' ratings of a specific course and instructor were chiefly a function of the teaching competence of that instructor and students' perception of his competence. V. de Wolf. (A7, C43, F11)

Feldman, K. A. Student assessment of teaching: A selected bibliography. Improving College and University Teaching, 1973, 21, 62-63.

More than 100 selected titles of articles on student assessment of teaching effectiveness are presented. Author. (A2)

Field, T. W., Simpkins, W. S., Browne, R. K., & Rich, P. Identifying patterns of teacher behavior from student evaluations. Journal of Applied Psychology, 1971, 55, 466-469.

A number of studies have attempted to identify the basic dimensions of variability underlying students' perceptions of teachers. It is possible that these perceptions reflect stylistic differences among teachers. Applying the method of discriminant analysis, it was shown that distinctive patterns of items were characteristic of students' discriminations. These patterns were interpreted as representing important configurations of teaching behavior, and the combination of behaviors for each teacher was viewed as a description of his personal teaching style. The teachers in the study were clearly differentiated in terms of these patterns of teaching behavior. Authors. (A6, A8, B24, I1, I4)

Finkbeiner, C. T., Lathrop, J. S., & Schuerger, J. M. Course and instructor evaluation: Some dimensions of a questionnaire. Journal of Educational Psychology, 1973, 64, 159-163.

This study attempted to discover the underlying dimensions involved in student attitudes toward courses and instructors. A course and instructor evaluation questionnaire was administered to 1,616 subjects at the academic centers and to 6,352 subjects at the main campus of Cleveland State University. The data were factor analyzed and yielded five rotated factors in each group, accounting for approximately 50% of the total variance. The factor matrices for the two groups were found to be significantly congruent. The five factors were interpreted as General Course Attitude, Attitude toward Examinations, Attitude toward Method, Instructor/Student Rapport, and Attitude toward Work Load. A multifactor model of course attitudes was supported. Authors. (A6, A9, B25, C14)

Follman, J. Modes and formats in student ratings scales of teaching effectiveness. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, April 1974.

The purpose of this study was to investigate the effects of kinds of response modes and kinds of rating formats on level of student ratings of teacher and course effectiveness. The specific response modes investigated were bipolar vs. unipolar rating continua. The kinds of formats used were: Agreement; Evaluation; and Needs Improvement. All formats were on a five-point scale: 5, 4, 3, 2, 1. Analysis of variance for the total mean score indicated that there were significant differences for formats ($p < .005$). These differences reflected the higher means, for both instructors being evaluated, associated with the Needs Improvement key vis a vis the Agreement and Evaluations keys. T. de Wolf. (A3, A11, B33)

Fox, A. M., & Hein, D. D. Teacher perceptions in effective teaching. *Colorado Journal of Educational Research*, 1971, *10*(3), 45-49.

The purpose of this study was to investigate the following questions: (1) can faculty members select from their class those students who would give high evaluations of their teaching effectiveness and those students who would give low evaluations of their teaching effectiveness as indicated by the University of Northern Colorado (UNC) Rating Scale? (2) Will the faculty members be able to predict the class evaluations of his/her relative strengths and weaknesses as indicated by the UNC Rating Scale? Of the sample investigated 74% of the faculty members were able to successfully predict their relative strengths and weaknesses in a particular group while only 26% of this same group successfully predicted which students would give them relatively high and low ratings. V. de Wolf. (B7, C49, H1, K2)

Foy, J. M. A note on lecturer evaluation by students. *Universities Quarterly*, 1969, *23*, 345-349.

One lecturer at a British university was assessed for lecturing technique in 1966 and 1968 by two entirely separate groups of students. Both groups used the same 43-item rating form. The students also rated their judgment of the relative importance of these particular characteristics for an ideal lecturer. The 1968 group of students rated the characteristics of an ideal lecturer slightly higher than their predecessors, although the correlation between the 1966 and 1968 values was r = .93. The values for the actual ratings of the lecturer correlated, as one might expect, somewhat less highly (r = .71). V. de Wolf. (I4, J3)

Frey, P. W. Comparative judgment scaling of student course ratings. *American Educational Research Journal*, 1973, *10*, 149-154. (ERIC: EJ076186) (a)

The present report examined a novel procedure for providing student instructional ratings for faculty evaluations. Four large introductory courses taught at Northwestern University, having considerable enrollment overlap, were sought. The courses selected were introductory classes in biology, chemistry, English, and history--each having a different instructor. Students who had taken two or more of these courses completed a 15-item questionnaire for each course they had taken. Factor analysis revealed 4 primary underlying factors. The psychological distance among the four courses on each of the four primary factors was estimated using case V of Thurstone's law of comparative judgments. The results of this scaling exercise indicated that the students reliably distinguished among the four courses on each of the four factors. The author discussed these results and concluded with a summary of the limitations of this procedure. V. de Wolf. (A6, B26, D10, G3)

Frey, P. W. Student ratings of teaching: Validity of several rating factors. *Science*, 1973, *182*, 83-85. (b)

Students' ratings of their instructors in undergraduate classes in calculus were correlated with class performance on a common final examination.

Ratings on several instructional factors were highly related to class performance even though they appeared to be independent of students' own grades. Author. (A6, A11, F25)

Gadzella, B. M. College students' views and ratings of an ideal professor. College and University, 1968, 44, 89-96. (a)

446 randomly selected students at Western Washington State College (WWSC) completed a 25-item questionnaire, Students' Views of an Ideal Professor. The results of the survey indicated that the five most important criteria in describing an 'ideal' professor were: mastery in knowledge of subject, deep interest in subject, great flexibility, well organized course preparation, and used appropriate vocabulary. V. de Wolf. (B2, C15, F2, F5, F6, F7, F30, F31, I4)

Gadzella, B. M. The ideal professor as viewed by New Paltz freshmen. Report No. 9. New Paltz, New York: Office of Institutional Research, State University College, 1968. (b)

Gadzella's questionnaire, Students' Views of an 'Ideal' Professor, was administered to 1145 freshmen at the State University of New York College, New Paltz. Students were asked to select the 5 criteria which in their opinion are 'Most Important' characteristics of an ideal professor and to indicate the relative importance of these selected criteria by arranging them on a 5-point scale. Similarly, they were then asked to select the 5 criteria which they considered to be the 'Least Important' characteristics of an ideal professor. Responses were analyzed for the total group, sex, intended curriculum, and academic major. The author thoroughly discussed these responses. Freshmen at New Paltz evidently saw mastery of subject matter and deep interest in and enthusiasm for the subject as the most important traits of an ideal professor. Whereas, appearance and ableness as an author were the two characteristics considered the least important by this group. Differences among these students by sex, curriculum, and major were not great. V. de Wolf. (B34, C15, F6, F7, I4)

Gates, M., & Burnett, C. W. Students' perceptions of professors. Improving College and University Teaching, 1969, 17, 234-236.

Utilizing Q sort, the present study attempted to determine if a difference existed between graduate and undergraduate students' perceptions of their professors. The correlation obtained between the group of undergraduate students and graduate students was -.69. The negative correlation of -.69 was significant at the .01 level. Thus, the results of this particular Q sort suggested that statements about professors which graduate students perceived as most like their professors, undergraduates perceived as least like their professors. V. de Wolf. (B35, F2)

Gessner, P. K. Evaluation of instruction. Science, 1973, 180, 566-570.

A sophomore class of medical students taking a one-semester basic science course were the subjects in this study. Student ratings of teaching effectiveness were correlated with class performance in 20 subject areas on

Part I of the National Medical Board Examination and also with class performance in the same subject areas on department examinations. It was concluded that student ratings of instruction and class performance on national normative examinations are positively related: the higher the student ratings of instruction, the higher the class score relative to a nationwide norm. On the other hand, no significant correlation was found between student ratings and class performance on institutional examinations. This suggested that both student ratings and class performance on national normative examinations were valid measures of teaching effectiveness. V. de Wolf. (F25)

Gillmore, G. M. Approximating decile norms for the Illinois Course Evaluation Questionnaire by use of the normal curve. Research Report #432. Urbana, Illinois: Measurement and Research Division, Office of Instructional Resources, University of Illinois, 1972. (a)

The purpose of this report is to compare two methods of deriving decile norms for the Illinois Course Evaluation Questionnaire. The definitional method uses the cumulative frequency distributions of the results of previous users. The normal approximation method uses the normal curve. The advantages of the two methods are discussed, and the norms generated by each method are compared for 5346 sections nation-wide, and for 2784 sections at the University of Illinois. The report suggests use of the normal approximation method. Author. (A13, B3, C5)

Gillmore, G. M. Three functions of student course evaluations. In L. M. Aleamoni (Chair), Course and instructor evaluation. Symposium presented at the National Council on Measurement in Education Convention, Chicago, April 1972. (b)

The normative, diagnostic, and informative functions of the Illinois Course Evaluation Questionnaire were discussed. V. de Wolf. (A11, B3, B12, C5)

Gillmore, G. M. Estimates of reliability coefficients for items and subscales of the Illinois Course Evaluation Questionnaire. Research Report #341. Urbana, Illinois: Measurement and Research Division, Office of Instructional Resources, University of Illinois, 1973. (a)

The major focus of this paper is on the reliability of the individual items, the subscales, and the total score of Form 66 of the Illinois Course Evaluation Questionnaire (CEQ).* The CEQ is a Likert-type attitude questionnaire designed to elicit evaluative information from students about courses in which they are enrolled. Form 66 contains 50 items, which are combined by unweighted averaging to form six subscales and a total score.

"Stability" coefficients were estimated by three methods applied to six samples. All estimates indicated reasonably high reliabilities for classes. The magnitude of these reliabilities were discussed in relation to number of students in a class.

"Equivalence" coefficients were presented for all subscales and discussed. Finally, reliabilities were discussed in the context of standard errors of measurement which was discussed, in turn, within the context of norming. Author. (A11, B3, B12, C5)

*The results of this paper can be used with CEQ Form 72 and Form 73 as it contains items and subscales found in Form 66.

Gillmore, G. M. Evaluation by students for university-wide comparative purposes. In H. R. Smock (Chair), A plan for the comprehensive evaluation of college teaching. Symposium presented at the American Educational Research Association Convention, New Orleans, February 1973. (b)

The development of an instrument to evaluate instruction was discussed in terms of its content domain, reliability, face validity and university-wide applicability. Data from the University of Illinois Course Evaluation Questionnaire was used to support the author's opinions. V. de Wolf. (A11, A14, B3, C5)

Gillmore, G. M., & Brandenburg, D. C. Would the proportion of students taking a class as a requirement affect student ratings of the course? Research Report #347. Urbana, Illinois: Measurement and Research Division, Office of Instructional Resources, University of Illinois, 1974.

CEQ-The Advisor ratings from 400 of 763 class sections from the spring of 1971 administration were divided into two groups of 200 sections each. The sections were separated into five groups depending on the proportion of students taking the class as a requirement. These proportions were as follows: more than 4/5 of the student enrollment was mandatory, 3/5 to 4/5 of student enrollment was mandatory, etc.

Identical analyses were performed on each sample with identical results. One-way analyses of variance (ANOVA) on each of the six subscales of the CEQ yielded statistically significant differences among the mean ratings of the five groups. Polynomial trend analyses subsequently showed that the linear trend was significant ($p < .01$) in each case. This result implied that as the proportion of students taking a class because of a requirement increases, the ratings given to the course and instructor decreases.

Suggestions were made to caution instructors about interpreting CEQ results for classes which have high or low proportions of requirement enrolled students. It was suggested also that further investigations be performed in this area before the results presented here are regarded as conclusive evidence. Authors. (C5, G5)

Good, K. C. Similarity of student and instructor attitudes and student attitudes toward instructor. Dissertation Abstracts International, 1972, 32(8-A), 4418-4419. (Ann Arbor, Michigan: University Microfilms. Order No. 72-7962.)

The primary purpose of this study was to investigate the relationship between actual and assumed student-instructor attitudinal similarity and the student's liking for or attraction to the instructor. In addition, the study was concerned with the possible relationship between student-instructor attitudinal similarity and the student's rating of the instructor on other evaluative dimensions such as competence and open-mindedness. Finally, the study investigated the relationship between student-instructor similarity and the student's rating of the course on aspects other than the instructor. Results of the study revealed a rather consistent pattern of significant correlations between assumed similarity and attraction. Several significant correlations between assumed similarity and other ratings of the instructor and course were also indicated. However, significant correlations between actual similarity and such evaluations were relatively infrequent. V. de Wolf. (D12, D18, E2)

Good, K. C., & Good, L. R. Attitude similarity and attraction to an instructor. *Psychological Reports*, 1973, *33*, 335-337.

It was hypothesized that a college instructor who is attitudinally similar to oneself will be evaluated more positively than an attitudinally dissimilar one for open-mindedness, promoting feelings of at ease, being stimulating and interesting, over-all teaching competence, personal attractiveness, and desirability as an instructor. Each undergraduate S filled out a 14-item Survey of Attitudes, and, during a later class session, received an attitude survey representing the attitudes and opinions of a hypothetical college instructor who showed either 14% or 86% aggreement with S's own views. S then filled out an Instructor Evaluation Scale for his evaluations of the stimulus person. The hypothesized effect of attitude similarity was confirmed for all of the evaluation variables. Authors. (C16, D1, E2, F6)

Graham, M. H. The relationship between CEQ ratings and instructor's rank, class size, and course level. Research Report #337. Urbana, Illinois: Measurement and Research Division, Office of Instructional Resources, University of Illinois, 1972. (ERIC: ED076147)

A study was conducted to determine if the tendency for faculty members of higher rank to receive the highest ratings on the Illinois Course Evaluation Questionnaire (CEQ) remained when variables such as class size and course level were taken into account. The relationship between CEQ ratings and instructor's rank, class size, and level of course was examined by means of multivariate analysis of variance (MANOVA). Dependent variables were the six subscales of the CEQ. As hypothesized there were no significant differences in ratings assigned by students in small (1-20 students), medium (21-40 students), and large (over 40 students) classes ($p < .13$), or received by teaching assistants, instructors, and assistant, associate, and full professors ($p < .13$). Highly significant differences ($p < .0001$), however, were found in ratings assigned by students in 100-, 200-, 300-, and 400- and 500-level courses. In addition, significant size by level ($p < .03$) and size by rank ($p \leq .02$) interaction effects were found. Discriminant functions computed for effects found to be significant yielded information concerning the extent and direction of these significant differences. Author. (A8, B3, C5, D4, G1, G4)

Grant, C. W. Faculty allocation of effort and student course evaluations. *The Journal of Educational Research*, 1971, *64*, 405-410. (ERIC: EJ041317)

Two types of data were obtained on teaching faculty at the University of Utah, student evaluations of courses, and faculty allocations of time distributed among academic activities. The data were subjected to treatment by analysis of variance procedures. It was found that faculty rank was not related to student evaluations of courses on any of the five course evaluation scales and in general there was little relationship between faculty allocations of time and student evaluations of courses. It was noted, however, that students in courses with large enrollments considered faculty more "prepared" than students in courses with smaller enrollments. There was some evidence that as faculty time allocated to research and writing increased, student ratings of courses decreased. Author. (B36, D4, D6, D15, G1, G4)

Granzin, K. L., & Painter, J. J. A new explanation for students' course evaluation tendencies. *American Educational Research Journal*, 1973, *10*, 115-124.

This study of the student course evaluation process discovered significant correlations between course ratings and variables representing commitment and course-end attitudes toward the course. It found relationships of lesser significance for attitude change measures, while demographics provided generally nonsignificant correlations. Stepwise regression equations developed for their power to predict course ratings relied most heavily on course-end attitude variables. Factor analysis of the variable set revealed 6 factors underlying the course evaluation structure studied, and this analysis guided formulation of new regression equations having reduced predictive power but greater independence among included predictor variables. Conclusions focused on the study's contributions to understanding the course evaluation process and suggested steps an instructor might take to improve his ratings. Authors. (A6, B36, E4, F1, F2, F3, F5, F6, F11, F23, F26, F27, F32, F33, F34, F36, F37, F38, G5, J5)

Greenwood, G. E., Bridges, C. M., Jr., Ware, W. B., & McLean J. E. Student Evaluation of College Teaching Behaviors instrument: A factor analysis. *The Journal of Higher Education*, 1973, *44*, 596-604.

The development of the Student Evaluation of College Teaching Behaviors (SECTB) was outlined. After factor analysis of the 60-item SECTB eight factors were identified for a large composite sample of student and faculty responses. V. de Wolf. (A6, A8, A9, A10, C17)

Gromisch, D. S., Bamford, J. C., Rous, S. N., Sall, S., & Rubin, S. A comparison of student and departmental chairman evaluations of teaching performance. *Journal of Medical Education*, 1972, *47*, 281-284.

The purpose of this paper was to compare the results of a student evaluation of the teaching performance of a medical school faculty with the results obtained when the same teachers were evaluated by their department chairmen utilizing the same questionnaire. The instructors were from three different departments. The findings of this study showed that students and chairmen did not rate faculty members similarly. The authors concluded that this indicated the need to define evaluation criteria better and to devise more precise methods by which to measure them. Until this was done there would be no improvement in learning experiences and hence in the quality of the product of medical education - the practicing physician. V. de Wolf. (H8)

Hall, D. T. The effect of teacher-student congruence upon student learning in college classes. Journal of Educational Psychology, 1970, 61, 205-213.

This research applies Pervin's concept of person-environment fit to learning in college classes. The aspect of the environment considered is the Teacher Style of the instructor, and the aspect of the person studied is his preferences in a teacher, or his Ideal Teacher Style. The hypothesis tested was that the fit between the teacher's actual teaching style and the student's ideal teacher style would correlate more highly with learning than would either of these constructs alone. This hypothesis was not confirmed; the actual teaching style of the instructor was generally a better predictor of learning than were the fit scores or the student's ratings of ideal instructor behavior. The differences between these negative findings and Pervin's positive findings are discussed in terms of (a) the variances, impact and validity of Actual and Ideal Teacher Style (or, more generally, of the environmental and personal constructs employed), (b) the nature of the particular environment studied, (c) the constructs measured, and (d) the instruments employed. These issues appear to have an important bearing upon the potency of the person-environment fit in different situations, relative to the potency of the person or the environment alone. Author. (A6, B38, D16, F3, F19, H1, I1, I4)

Hanke, J. E. Teacher and student perceptions as predictors of college teaching effectiveness. Dissertation Abstracts International, 1971, 31(8-A), 3778. (Ann Arbor, Michigan: University Microfilms. Order No. 71-4182.)

The problem investigated in this dissertation was to explore the areas of conceptual system orientation, perceptual orientation, and levels of learning as they were perceived by both teachers and their students in relation to student ratings of college teachers. Knowledge of the 20 predictor variables employed in this study provided moderately efficient prediction of college teacher effectiveness for all 13 criteria. The more effective college teachers were perceived as abstract-oriented people who gave the impression of complexity of thought and feeling. They were perceived as being capable of admitting new and different ideas or opinions, examining and entertaining them, and accepting or rejecting them on an objective basis. They were perceived as weighing the evidence in each situation and being tolerant of diverse opinions. They were open to all ideas and changes from every available source. Efficient college teachers perceived other people as basically able, worthy, dependable, and internally motivated and have good, positive self concepts. The experiences that they provided in their classes were designed to enable students not only to be able to explain and apply the material that was covered, but also to have the inclination and opportunity to do so. V. de Wolf. (B7, D1, D3, D10, D12, D19, D23, D24, D25, D26, F2, F19, G1, G5)

Hanke, J. E., & Houston, S. R. Teacher and student perceptions as predictors of college teaching effectiveness. College Student Journal, 1972, 6(1), 45-46.

This article was an expanded summary of Hanke's dissertation. See preceding abstract entitled "Teacher and student perceptions as predictors of college teaching effectiveness" by Hanke (1971). V. de Wolf.

Hanke, J. E., Houston, S. R., & Usher, R. Researching the effective college teacher: A perceptual approach. Cheney, Washington: Department of Accounting and Quantitative Business Analysis, Eastern Washington State College, no date. (Duplicated report)

This article was essentially identical to "Teacher and student perceptions as predictors of college teacher effectiveness" by J. E. Hanke and S. R. Houston. The only additions were a 3-paragraph introduction and a bibliography. Similarly, like the Hanke and Houston article, this duplicated report was a summary of Hanke's dissertation. See "Teacher and student perception as predictors of college teacher effectiveness" by Hanke (1971). V. de Wolf.

Harry, J., & Goldner, N. S. The null relationship between teaching and research. Sociology of Education, 1972, 45, 47-60. (ERIC: EJ051776)

Analysis of the data from a survey of members of a university faculty and of student attitudes toward this faculty indicate no relationship between student-rated adequacy of teaching and the extent of a teacher's scholarly activity. It appeared that student criteria for determining an instructor's adequacy consisted largely of evaluations of the instructor's style. Research activity by faculty was related to content variables such as course difficulty and student study time. The commonly-held view of a conflict between teaching and research found no support, although there may be some conflict when longer-run effects are explored. Authors. (A6, C18, D6, D8, D15)

Hartley, E. L., & Hogan, T. P. Some additional factors in student evaluations of courses. American Educational Research Journal, 1972, 9, 241-250.

The purpose of the present study was to change the focus of attention in course evaluation from the course-instructor complex to the student's estimate of self-development as a result of the course. Two instruments were used for the evaluation of courses: the first was a traditional-type evaluation of course content and teaching; and the second scale asked the student to rate his/her own self-development as it was affected by the course. Factor analysis of approximately 450 student questionnaires showed that there appeared to be little that was common to the two instruments used in this study. Four of the seven factors were defined exclusively by terms within the first questionnaire and the other three by items of the second scale. This clear separation of items in both scales suggested the need for continued exploration and expansion of instruments for securing student evaluations of courses. V. de Wolf. (A6, A9, B39)

Harvey, N., & Barker, D. G. Student evaluation of teaching effectiveness. Improving College and University Teaching, 1970, 28, 275-278.

Using the point biserial coefficient of correlation, ratings on each of the 21 items of this ratings scale were correlated with the criterion classification as "most effective" or "least effective" teacher. A total score on the instrument was also obtained by summating the ratings assigned each teacher rated, and this variable was also used as a criterion for evaluating the items of the rating form, using the product-moment coefficient of correlation. There was approximately the same pattern of correlations with general subjective evaluation as with summated score on the rating instrument. Two items, however, were distinctly less positively correlated than the other 19 items with both criteria: (1) instructor's involvement in research; and (2) general estimate of course. V. de Wolf. (A3, A10, B40, D6)

Hayes, J. R. Research, teaching, and faculty fate. Science, 1971, 172, 227-230. (ERIC: EJ038399)

The author tried to answer three questions in this article: (1) Are research activity and teaching ability related to each other? (2) In what ways do research activity and teaching ability influence classroom assignment? (3) In what ways do research activity and teaching ability influence promotion? The answer to the first question was not wholly clear. If one took department heads' judgments at face value, there was evidence of a strong positive relation between research ability and teaching quality. If, on the other hand, one interpreted the correlation in the department heads' judgments as a "halo" effect, then there was no evidence, in the six relations measured in this study, that research activity and teaching ability were related. Answers to the second and third questions appeared quite clear. Individuals with high research ability and high rank tended to be assigned high-level classes. Teaching quality was unrelated to classroom assignment. Promotion was strongly related to measures of research activity but appeared to be unrelated to teaching ability. V. de Wolf. (B41, D4, D6, D7, G4, H8)

Heien, L., & Moody, R. Current research on learning and teaching. Educational Perspectives, 1973, 12(4), 21-26.

The second section of this article outlined studies concerning the evaluation of teaching effectiveness as assessed by two approaches: effectiveness as the general impact on the learner; and effectiveness as achievement of specific goals. Under this second approach for defining teaching effectiveness, Moody related the results of an unpublished study of his own in which he used step-wise regression analysis to investigate teacher ratings and residual gain scores of students in first semester Spanish. The article concluded with a short review of studies which relate evaluation of teaching effectiveness to the improvement of teaching. V. de Wolf. (A2, D1, D12, D16, F2, F3, F6, F7, F25, H3, J3)

Hein, D. Effectiveness of teaching. Colorado Journal of Educational Research, 1971, 10(2), 18-22.

In this article, the author attempted to review the dissertations completed at the University of Northern Colorado since June, 1954 in the area of teacher effectiveness studies. Four such studies were included under the sub-grouping of college teacher effectiveness. Two of these were related to the student ratings of instruction. V. de Wolf. (A2, B7, D2, D3, D10, F6, F7, F12, F41, F61, F67, F68, F69, F70)

Heyn, D. R. Development and validation of a sociometric instructor evaluation instrument and procedure. Sherman, Texas: Austin College, 1972. (Duplicated report) (ERIC: ED068077)

Professors and courses were evaluated sociometrically by almost 90% of all full-time resident students. The three professors and courses they would most as well as least recommend to other students were listed. A 30% random sample of respondents was then interviewed by other students. Both the sociometric results and the results of the interviews were given to faculty residents during the summer (approximately 70%). They were asked to respond to a questionnaire primarily devised to indicate credibility and defensiveness. Half the faculty were given coded interview results and half verbatim results. No significant differences in response were found. Also, there was some evidence that the procedure produced changes in faculty attitudes toward themselves and their courses. Results were also scrutinized as to tenure or none, years at Austin College, rank, age, choice quartiles (high most-high least, high most-low least, low most-high least, and low most-low least), highest degree earned, and area (Humanities, Social Sciences, or Physical Science). Results were sensible and the null hypotheses disconfirmed. Results and implications were discussed. Author. (A9, B42, D3, D4, D8, D9, D10, D27, G6, J3)

Hicks, R. A. The relationship between publishing and teaching effectiveness of college professors. San Jose: San Jose State University, 1969. (Duplicated report)

A number of observers of higher education have been critical of the writing and research involvements that occupy the time of many college professors, claiming that hours spent in this manner detract from effective teaching. (For review, see Bresler, 1968). Recently, however Bresler (1968) has demonstrated that professors who engaged in grant-supported research were rated by students as more effective teachers than professors who had not been involved in grant-supported research. The purpose of this study was to extend Bresler's finding by determining whether a similar relationship existed between publishing and teaching effectiveness. Author. (B43, D6)

Hildebrand, M. The character and skills of the effective professor. The Journal of Higher Education, 1973, 44, 41-50.

This study had two objectives: first, to identify and describe effective teaching so that instructors could be helped to improve their teaching effectiveness; and second, to find more valid, reliable, effective means of incorporating the evaluation of teaching into advancement procedures. It was a summary of a full report entitled "Evaluating University Teaching," by M. Hildebrand, R. C. Wilson, and E. R. Dienst, an abstract of which is included in this bibliography. V. de Wolf. (A6, A9, A10, B44, D10, D14, D16, F1, F2, F7, F11, F39, G1, G4, H3)

Hildebrand, M., Wilson, R. C., & Dienst, E. R. Evaluating university teaching. Berkeley: Center for Research and Development in Higher Education, University of California, Berkeley, 1971.

Students and faculty were asked to identify and characterize the best and worst instructors in the faculty. A great deal of agreement existed between the two groups in identifying effective and ineffective teachers. Factor analysis of the items characterizing best teachers as perceived by students produced 5 factors. In general, student ratings of best teachers showed only negligible correlations with academic rank of instructor, class level, number of courses previously taken in the same department, class size, required versus optional course, course in major or not, sex of respondent, class level of respondent, grade-point average, and expected grade in course. A disproportionate number of best teachers were teaching seminar rather than lecture courses, and a wide range of excellence was revealed in the teaching of different subject areas. V. de Wolf. (A3, A6, A9, A10, A11, B44, C38, D4, D6, D10, D14, D15, D16, E4, F2, F5, F6, F9, F23, F39, G1, G4, G5, H3)

Holmes, D. S. The relationship between expected grades and students' evaluations of their instructors. Educational and Psychological Measurement, 1971, 31, 951-957. (a)

The Teaching Assessment Blank (TAB) was used in the College of Arts and Sciences at the University of Texas (Austin) to assess the relationship between expected grades and student evaluation of instruction. It was found that, in most classes, relationships existed between the grades students expected and the degree to which they reported that they were stimulated by the instructor and the degree to which they felt that the grading system was fair. On the other hand, items assessing the instructors' presentations were not found to be related to the expected grades. While the nature and antecedents of the observed relationships were of considerable theoretical interest in terms of both the evaluation process and the possible differential effect of instructors in different types of students, it was clear that the relationships did not severely distort the evaluation process and therefore did not pose a serious threat to the validity of the evaluation system. V. de Wolf. (B21, C19, F23)

Holmes, D. S. The teaching Assessment Blank: A form for the student assessment of college instructors. The Journal of Experimental Education, 1971, 39(3), 34-39. (b)

A form which students can use to assess their class experiences was presented. A factor analysis based on evaluations filled out by 1,648 students revealed four factors which measured (a) the quality of the instructors' presentations, (b) the evaluation process and the student-instructor interactions, (c) the degree to which the students were stimulated and motivated by the instructors, and (d) the clarity of the tests. A further analysis indicated that subscale scores which reflected the factor scores could be developed from the total item pool. Author. (A6, A9, B21, D19)

Holmes, D. S. Effects of grades and disconfirmed grade expectancies on students' evaluations of their instructor. Journal of Educational Psychology, 1972, 63, 130-133.

Half of the students who deserved and expected A's or B's were given their expected grades, while half were given a grade one step lower than expected. After receiving the grades, the students filled out the Teaching Assessment Blank. A 2 X 2 analysis of variance revealed that the evaluation of the instructor was lowered on only 1 of the 19 items as a function of differences in grades, but evaluations on 10 of the 19 items were lowered as a function of the unexpected lowering of grades. It was concluded that although differences in actual grades do not affect evaluations, if students' grades disconfirm their expectancies, the students will tend to deprecate the instructor's teaching performance in areas other than his grading system. Author. (C19, F3, F23, F24)

Hoyt, D. P. The Kansas State University program for assessing and improving instructional effectiveness. In A. L. Sockloff (Ed.), Proceedings: The first invitational conference on faculty effectiveness as evaluated by students. Philadelphia, Pa.: Measurement and Research Center, Temple University, 1973. Pp. 152-163. (a)

A brief history of the teacher evaluation program at Kansas State University was given in this article. V. de Wolf. (A9, All, B45, J3, K3)

Hoyt, D. P. Measurement of instructional effectiveness. Research in Higher Education, 1973, 1, 367-378. (b)

The problem of measuring instructional effectiveness was examined, and a rationale was offered for employing "student progress on relevant objectives" for this purpose. To assess such progress, it was suggested that instructor ratings of the importance of objectives be combined with student ratings of progress on these objectives.

On the basis of this suggestion, data were collected from 708 undergraduate classes at Kansas State University. An analysis of these data resulted in the following conclusions:

1. Faculty members appeared to make reliable judgments of the relative importance of these objectives.
2. Student progress ratings were made with acceptable reliability when there were 20-25 raters. Reliability of the overall progress measures was satisfactory when only 10 raters were used.
3. Students used some discrimination in rating progress on various objectives, but their ratings were also noticeably subject to the halo effect.
4. An indirect test of the validity of class progress ratings yielded positive results.

The proposed method of evaluating instruction appears generally feasible and useful. Its application would provide a practical approach to judging teaching success. Such an approach is essential before investigations can be undertaken of how teaching might be improved. Author. (A9, A11, B45, D22, D28, G8)

Jandt, F. E. A new method of student evaluation of teaching. Improving College and University Teaching, 1973, 21, 15-16.

The evaluation described in this article compared students' evaluations of college courses in general (and by inference their expectations for a college course) with their subsequent evaluations of a particular course. Two new evaluation forms were constructed each having 20 five-interval rating items. One was designed to specifically cover major factors of the course and was administered at the beginning of the course. The second form was designed to parallel the first, but with references to the specific college course in general and was administered at the end of the course. Using t-tests for independent samples to determine significant differences between the means for each item, 18 of the 20 items yielded significant comparisons. The same 2 evaluation forms were used in a later quarter. Comparing results of one item from both of these administrations, the author concluded that different classes do hold different expectations for a college course and that an evaluation which does not account for the effects of expectations on the evaluative process may present distorted evaluation V. de Wolf. (B11, I1, I4)

Jiobu, R. M., & Pollis, C. A. Student evaluations of courses and instructors. The American Sociologist, 1971, 6, 317-321.

The purpose of the present study was to examine how a number of variables affect students' perceptions of courses and instructors. The data were collected from course evaluations at the University of Wisconsin-Green Bay (UWGB). 18 independent variables were used in a stepwise regression analysis in relation to the 2 dependent variables-- amount of student perceived learning and overall evaluation of course. The results showed two independent variables which might affect overall evaluation but not affect perceived learning. Courses in which upper-level students predominated were rated somewhat more favorably than courses in which lower-level students predominated, and classes containing students with poor average grades were slightly less favorably perceived than were classes that had students with higher grade averages. The data showed that class size did not affect evaluations and that expected grade was not related to either evaluation measure. V. de Wolf. (A9, A11, B39, F1, F2, F5, F6, F19, F23, G1)

Kane, M. T., Gillmore, G. M., & Crooks, T. J. The application of generalizability theory to course evaluation questionnaires. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, April 1974.

The major purpose of this paper was to present an organized and comprehensive view of the sources of variability in student ratings of instructional effectiveness, when these ratings were in the popular context of a uniform, multi-item, objective questionnaire applied to a number of distinct, intact classes. The analysis employed was directly derived from generalizability theory. Actual data from two instruments were presented, and several different course and instructor restrictions were employed for specific analysis. V. de Wolf. (A11, B3, C5, C55, D10)

Kennedy, W. R. The relationship of selected student characteristics to components of teacher/course evaluations among freshman English students at Kent State University. Dissertation Abstracts International, 1972, 32(9-A), 5038-5039. (Ann Arbor, Michigan: University Microfilms. Order No. 72-9270.)

This investigation of the relationship of selected student variables and the evaluative components of a teacher/course evaluation instrument was conducted using as a sample 549 freshman students enrolled in English 160 at Kent State University during Fall Quarter, 1970. Some of the findings were: (1) Students' age, Omnibus Personality Inventory subscale scores, American College Test scores, and areas of agreement with teachers concerning the qualities of the "best-teacher" and "best-student" were not correlated with evaluative subscales; (2) Students' Fall Quarter gpa, actual grade in the course, and "actual versus expected grade" differential were positively correlated with all evaluative subscales; and (3) Students' expected grade, college within the University, and sex did differentiate certain of the evaluative subscale scores. V. de Wolf. (B65, E5, F1, F3, F5, F6, F7, F12, F16, F23, F24)

King, A. P. The self-concept and self-actualization of university faculty in relation to student perceptions of effective teaching. Dissertation Abstracts International, 1972, 32(7-A), 3615. (Ann Arbor, Michigan: University Microfilms. Order No. 72-4758.)

From this study of university instructors rated high and low by students, and from limited psychological testing, the following conclusions were reached: (1) College affiliation and number of years of teaching seemed unrelated to teacher effectiveness, but sex, age, highest degree earned, and years of formal education made a difference in distinguishing between good and poor instructors; (2) Analysis of variance scores showed that there was no difference between mean scores of good and poor instructors in relation to how they accepted themselves and how they viewed their real self in relation to their ideal self-- both groups showed congruency and a positive view of self; and (3) When good and poor instructors were compared by an analysis of variance as to their self-actualization, there was generally no difference between the means-- generally poor instructors were just as self-actualizing (i.e., releasing of full capabilities and potentialities) as good instructors. V. de Wolf. (C5, D1, D2, D3, D4, D9, D10, D12)

Kohlan, R. G. A comparison of faculty evaluations early and late in the course. The Journal of Higher Education, 1973, 44, 587-595.

The present study was conducted at the University of Nebraska at Omaha to learn more about the stability of the faculty evaluation process. The primary hypothesis that student evaluations of their instructors would be quite stable after the first few hours of class, i.e., evaluations made early and late in the course would be highly associated, was supported. A secondary hypothesis that a dependent variable evaluating instructor characteristics like friendliness and respect for students would be more stable than other dependent variables was also supported. Females gave significantly higher ratings to instructors only on dependent measure reflecting structure and overall performance. Upperclassmen rated instructors more favorably than underclassmen. And lastly, the current results showed a curvilinear relationship between class size and stability of ratings. V. de Wolf. (A6, B46, C21, D28, F2, F5, F6, G1, J5)

Kooker, E. W. The relationship of known college grades to course ratings on student selected items. The Journal of Psychology, 1968, 69, 209-215.

Items for a rating scale were derived from college students' descriptions of what they liked or disliked about their college courses. These items were organized into seven areas. Students were asked to rate and re-rate a class they had had the previous semester. Test-retest reliabilities for the sections of the scale fell, with one exception, between .73 and .87. Total score reliability was .91. None of the test-retest means differed significantly. Samples of upper-class and freshman students were divided into three groups based on the grades they knew they had received--A, B, C and below. "A" students gave significantly more positive ratings than "C" groups in all areas except for "Teacher Presentation" and "Textbook" in the upper-class sample. Author. (A9, A11, F2, F3)

Krupka, J. G. Report on faculty and student evaluation of Instructor Rating Questionnaire. Northampton, Pa.: Northampton County Area Community College, 1970. (Duplicated report) (ERIC: ED038964)

Faculty and students at Northampton County Area Community College participated in the evaluation of an Instructor Rating Questionnaire (IRQ). This report compared the manner in which the faculty and students rated the twelve areas of the IRQ and the scores that they thought represented good and poor rating for each area. A copy of the questionnaire was attached. Author. (A10, B47, C22, I1)

Lancaster, O. E. Methods of faculty evaluation and development. Engineering Education, 1971, 62, 95-97. (ERIC: EJ054516)

The College of Engineering at Pennsylvania State University contacted students of various professors in the department in order to collect information after students had had time to use the material learned. A random sample of 40 students was selected for each faculty member from rosters of students taught by them two years before. A six-item questionnaire was used. The author concluded that "it can be shown that these student opinions are reliable and seem to be independent of the grades received." V. de Wolf. (B10, J5)

Leonard, W. M., II. Student preferences for what makes a good college teacher. Improving College and University Teaching, 1973, 21, 10-13.

The research reported herein attempted to tackle student preferences in classroom instruction. A sample of students at Illinois State University was confronted with a series of 14 forced choice questions. Chi-square analysis was done on each set of responses. Each of the 14 questions was then cross-tabulated with three variables: sex of respondent, over-all grade-point average (trichotomized into upper, middle, and lowest third), and plans in regard to graduate school (trichotomized into definitely intend to go, thinking of going, and do not intend to go). Once again chi-square analysis was done. Results of these analyses were summarized and discussed. V. de Wolf. (B48, F5, F6, F15, I4)

Levinthal, C. F., Lansky, L. M., & Andrews, O. E. Student evaluations of teacher behaviors as estimations of real-ideal discrepancies: A critique of teacher rating methods. Journal of Educational Psychology, 1971, 62, 104-109.

Ratings imply comparative judgments between the values of the observer and his observations, for instance, a student's ratings of his teacher are estimates of the discrepancy between the student's ideals for the teacher's behavior and what he sees the teacher do. However, most methods for collecting teacher ratings make assumptions about ideals and about the discrepancies between ideals and observed behavior. In order to assess the relevance of direct measurement of ideals, students were asked to rate a teacher and to report their ideals for the teacher's behaviors. The judgments of ideal behavior varied across students and items; there were also interactions between ideals and observed responses. Thus, the authors recommend new approaches to teacher ratings. Authors. (B49, F6, I1, I4)

Lewis, L. S. Students' images of professors. The Educational Forum, 1968, 32, 185-190.

Professors in three academic areas (science, social science, and humanities) and two nonacademic occupations (medical doctor and banker) were rated by students at a Northeastern state university. The students utilized a rating instrument consisting of 20 paired adjectives. The means of the academicians were compared alone, then the academician means were looked at in relation to the nonacademician means. Next, an analysis was made to determine whether students with dissimilar social or academic backgrounds differed on the images they held of academicians. No consistent differences were found among students when they were divided by sex, social class, year in college, grade point average, or extent of extracurricular participation, while marked differences were found among those in different majors. V. de Wolf. (C24, D10, F2, F5, F6, F7, F13, F40, I4, K5)

Lucas, J. A. Learning efficiency of students in varying environments. The Journal of Experimental Education, 1970, 39(1), 63-68. (a)

In determining whether or not students learned more effectively in different learning environments, experimental and control classes were conducted over a 2-year period in a basic statistics course that involved ten faculty members and 1,172 students. Grades, common examinations, and a checklist were the basic criterion measures used; mathematics background an and overall grade point (GPA) served as control mechanisms. Major conclusions were: students preferring some type of independent study consistently underachieved; instructor types, which produced high achievement levels for specific types of students, were identified in behavioral terms; it probably will take drastic manipulations of the learning environment beyond the normal constraints of the university in order to produce effective changes in the learning pattern. Author. (B50, C23, D16, F1, F3, F7, F25, F41, F42, K3)

Lucas, J. A. Optimal learning environments for different types of students. Dissertation Abstracts International, 1970, 30(11-A), 4877. (Ann Arbor, Michigan: University Microfilms. Order No. 70-7585.) (b)

This project considered the question of whether or not different types of students required different learning environments in order to learn most effectively. To answer this question comparisons and experimental classes were conducted over a two year period in a basic statistics course taught by ten different instructors and involving 1,172 students. Grades and common examinations corrected for entering differences in grade point and mathematics background were the basic criterion measures. An evaluation questionnaire and the Teacher Performance Checklist were the other criterion measures used.

The three major conclusions were as follows: (1) Students preferring some type of independent study consistently underachieved. (2) Instructor types, which produced high achievement levels for older students and marketing majors, were identified in behavioral terms. (3) It probably will take drastic manipulations of the learning environment beyond the normal constraints of the university in order to produce effective changes in the learning pattern. Author. (C23, F1 F3, F7, F16, F25, F41)

Lunney, G. H. Attitudes of senior students from a small liberal arts
college concerning faculty and course evaluation. Research Report #29.
Danville, Kentucky: Office of Institutional Research and Evaluation,
Centre College of Kentucky, 1973.

A study was conducted in an attempt to focus attention on the attitudes with which students approach the rating of instruction and to provide data concerning those attitudes. Ninety-seven percent (145 students) of the senior class of Centre College responded to a 37-item questionnaire concerning several aspects of faculty evaluation. The seniors felt that they could and did evaluate faculty fairly on the basis of valid criteria but they wanted the responses to go back to the instructor for his use and did not want them to be weighted heavily for administrative decisions concerning salary increases, promotion and tenure. An analysis of the responses by academic orientation (Humanities, Social Studies, and Sciences) and by sex shows that attitudes differed on several questions. Humanities majors tended to be unsure of their attitudes and to be more idealistic in their responses. Science majors seemed to be more pragmatic in their responses. Author. (B51, F6, F7, J3)

Magoon, J., & Bausell, R. B. Required versus elective course ratings.
College Student Journal, 1973, 7(1, part 1), 29-33.

This study was conducted in order to determine whether general student ratings of college instructors and courses are influenced by whether the courses being rated are required or elective. A sample of 2400 student ratings on a 29-item rating form were analyzed via a multivariate analysis of variance, followed by a multiple-group discriminant analysis of required versus elective group differences. The influence of required versus elective courses on the ratings is generally small, though statistically significant. It is suggested that this potential bias be viewed as relatively unimportant by local consumers of such information. Authors. (A7, A8, B15, F23, G5, K2)

Majer, K., & Stayrook, N. Reliability of college classroom course
evaluations. Paper presented at the annual meeting of the National
Council on Measurement in Education, Chicago, April 1974.

Two course evaluation questionnaires used at Indiana University were compared for stability and for their abilities to discriminate between effective and ineffective teachers. The method of measuring stability in this study was to repeat two of the items within each instrument-- thereby creating a "lie scale." Only about 1/3 of the students consistently rated their instructors the same on these identical items. The second half of this study examined the discriminating ability of the instruments with analysis of variance and intraclass correlation procedures. Internal consistency was confirmed in this study by the significant F ratios that were calculated when computing the intraclass correlations. However, the absolute value of the F ratios and the subsequent low intraclass correlation coefficients, suggested, since the variance in ratings within the classrooms was sufficiently large, that differences between teacher ratings cannot be easily attributed to the individual teacher. The results of this investigation suggested there were serious questions about the reliability of the two most widely used student response course evaluations on the Indiana University Bloomington Campus. V. de Wolf. (A4, A11, B9, C51, C52)

Mann, W. R. Changes in the level of attitude sophistication of college students as a measure of teacher effectiveness. Dissertation Abstracts, 1969, 29(8-A), 2443-2444. (Ann Arbor, Michigan: University Microfilms. Order No. 69-2351.)

The purpose of this study was to examine the instructor's behavior in the classroom and the personality and achievement characteristics of college students as possible variables to explain changes in attitude sophistication. The research design included an experimental group of 286 students in an introductory economics course and a control group of 31 history students. The major, statistically significant results of this study were: (1) Since the experimental group increased in attitude sophistication more than the control group, the course content and the instructors did make a difference; (2) The course grade was positively related to Achiever Personality, the student's SAT math score, a test of economic understanding taken before the course began, a reading accuracy test, and General Teaching Skill, a factor derived from a student evaluation of instructor form; (3) Attitude sophistication was positively related to Social Science Interest, the student's SAT verbal score, Student-Teacher Rapport, and Change in Beliefs-- the last 2 being factors from the student evaluation of instructor form; and (4) Attitude sophistication and the course grade were determined by different sets of instructor and student variables. V. de Wolf. (F3, F11, F12, F16)

McClelland, J. N. The effect of student evaluations of college instruction upon subsequent evaluations. California Journal of Educational Research, 1970, 21, 88-95.

It was the purpose of this study to test the hypothesis that student evaluation of college instruction would be affected by publishing the results of former evaluations. One week before final examinations, students in 6 classes rated various aspects of their courses on 18 nine-point scales. Students in two of the classes received questionnaires on which there was no indication of what former students had said about the course. However, students in the 4 remaining classes received questionnaires on which had been over-printed in another color the course number, the name of the instructor, and an 'x' in one box for each item. The students were told that these x's represented, for each item, the average response of students in all of the classes the instructor had taught in that subject during the previous semester and that these marks were to be ignored in making their decisions about the present course. Despite instructions to ignore the overprinted ratings, students consistently marked their ratings in the direction they were manipulated. Thus, tentative acceptance of the hypothesis that student opinion could be affected by publishing results of earlier student ratings was justified. V. de Wolf. (A7, B52, J1)

McDaniel, E. D., & Feldhusen, J. F. Relationships between faculty ratings and indexes of service and scholarship. Proceedings, 78th Annual Convention, APA, 1970, 619-620.

Several questions were asked about the relationships between effective teaching, publications, and research activities of college instructors. An 18-item instructor rating instrument, developed by the Purdue Course and Instructor Evaluation Foundation, was used. The research results indicated that the most effective instructors were those who write no books and those who limit their roles as paper and article writers to second authorship. Other results showed no relationships between research activity, as indicated by grants received, and instructional effectiveness. Time spent counseling students or supervising laboratories was found to be positively related to instructional effectiveness while time spent lecturing or in administration was negatively related to instructional effectiveness. Lastly, the results clearly indicated that the larger the class the lower was the ratings of instructional effectiveness. V. de Wolf. (D6, D7, D11, D15, G1)

McGlone, E. L., & Anderson, L. J. The dimensions of teacher credibility. The Speech Teacher, 1973, 22, 196-200.

In speech communication source credibility has been defined as the effects of receiver evaluations of the sources of messages on the impact of these messages. This current project was designed to identify the dimensions of teacher credibility, to assess their stability over time during a complete school term, and to develop appropriate instruments for measuring these dimensions. The instrument used consisted of 54 scale items from semantic differential scales used in recent credibility studies. The questionnaires were administered 4 times over a ten-week period: weeks 1, 3, 5, and during final examination period. Four factor analyses were obtained, corresponding to the four times at which the questionnaires were administered. At any given time there were major and minor factors operating which influenced student evaluations of teacher credibility; nevertheless, this factor structure was not stable over a period of time. The initial and final factor analyses produced relatively unified and discrete, if substantially different, factor structures. Over this period of time, a change was observed in concern from matters relating to the expertness of the teacher to those relating to his/her personality. The most certain conclusion reached by the authors was that the traditional categories of source credibility should not be used to delineate experimental treatments or to construct instruments for investigating teacher credibility. V. de Wolf. (A6, B67, D29)

McKeachie, W. J. Student ratings of faculty. AAUP Bulletin, 1969, 55, 439-444. (ERIC: EJ013328)

Because of the growing interest in making more effective use of student evaluations of courses and teachers in higher education, this article was prepared in hopes of offering suggestions and comments regarding the utilization of such evaluative procedures. A review of research in student ratings was included, as was a sample evaluation form. V. de Wolf. (A2, B8, B19, B53, B54, D1, D4, D9, D12, D14, F1, F2, F3, F6, F8, F16, G1, G5, H5)

McKeachie, W. J. Research on college teaching. Memo to the Faculty No. 44. Ann Arbor, Michigan: The Center for Research on Learning and Teaching, 1971.

The author reviewed research studies which have been conducted on the topic of college teaching. V. de Wolf. (A2, B10, B17, B19, B20, B22, B56, B57, B58, B59, D16, F6, F11, F12, F16, F41, G1)

McKeachie, W. J. Student evaluations keyed to function. Paper presented at the meeting of the American Educational Research Association, Chicago, April 1972.

The development and analysis of a form for student ratings of college teachers were explained. A multiple discriminant analysis of the student ratings yielded 10 significant dimensions. These dimensions were thoroughly described. V. de Wolf. (A6, A8, A9)

McKeachie, W. J. Correlates of student ratings. In A. L. Sockloff (Ed.), *Proceedings: The first invitational conference on faculty effectiveness as evaluated by students*. Philadelphia, Pa.: Measurement and Research Center, Temple University, 1973. Pp. 213-218.

The author briefly reviewed correlates of student ratings and divided these correlates into three areas: student characteristics; course or class characteristics; and characteristics of the instructor. Although McKeachie mentioned several sources in the body of the text, he failed to include any complete references in the text or a bibliography after the article. V. de Wolf. (D1, D2, D3, D4, D6, D9, D11, D12, D14, D18, D19, F1, F2, F3, F6, F8, F12, G1, G5)

McKeachie, W. J., & Lin, Y. Sex differences in student response to college teachers: Teacher warmth and teacher sex. *American Educational Research Journal*, 1971, *8*, 221-226.

The relationships between student achievement, student sex, instructor warmth, and instructor sex were examined in this study. The authors concluded that "with male teachers, high teacher warmth results in relatively high achievement for women students but not for men. For women instructors, high warmth teachers seem to be more effective with both sexes of students than teachers with low warmth." V. De Wolf. (D1, F3, F6)

McKeachie, W. J., Lin, Y., & Mann, W. Student ratings of teacher effectiveness: Validity studies. *American Educational Research Journal*, 1971, *8*, 435-445.

The authors reported the results of 5 of their studies which examined the validity of the student ratings in terms of criteria of student achievement.

Their results did not invalidate the use of student ratings as one source of evidence about teacher effectiveness, but they were less convincing than the authors had hoped for. Some major findings were: (1) In 4 of the 5 studies teachers rated high on the "Skill" factor tended to be effective with women students; (2) In all 5 studies teachers rated high in the "Structure" factor tended to be more effective with women than with men. In fact, on the whole, the more structured instructors tended to be ineffective for male students; (3) Teachers who were high in the "Rapport" ("Warmth") factor tended to be effective on measures of student thinking; and (4) Teachers whom students rated as having an impact on beliefs were effective in changing attitudes. V. de Wolf. (C8, F3, F6, F11, F25)

Menard, T. L. An analysis of the relationship between teacher effectiveness and teacher appearance. Dissertation Abstracts International, 1973, 33(7-A), 3394-3395. (Ann Arbor, Michigan: University Microfilms. Order No. 73-294.)

The purpose of this study was to determine if a relationship existed between teacher effectiveness and teacher appearance as judged by student ratings and also to determine if the student characteristics of sex, major, achievement, and socioeconomic status were predictors of teacher effectiveness based upon a difference in teacher appearance. Identical subject matter, methods of instruction, and tests were used in the 4 introductory psychology courses taught by the author in 2 consecutive quarters. The only difference was in the appearance of the instructor. Expert raters, using the "Self as an Instrument Scale," determined that there was no significant change in the author's personality resulting from the change in appearance. It was concluded that there was no difference in teacher effectiveness as measured by student ratings or student gains regardless of the appearance of the teacher and that the student characteristics of sex, major, achievement, and socioeconomic status did not aid in the prediction of teacher effectiveness. V. de Wolf. (B7, C44, D12, D35, F6, F7, F13, F16, F25)

Meredith, G. M. Dimensions of faculty-course evaluation. The Journal of Psychology, 1969, 73, 27-32.

The study attempted to establish the dimensionality of faculty-course evaluation. The Illinois Course Evaluation Questionnaire and A Student's Rating Scale of an Instructor was administered to 1097 students at the University of Hawaii. Sixty-seven variables were intercorrelated and factor analyzed, resulting in a nine factor solution. Two of the factors, labeled instructor impact and instructional impact, accounted for 64 percent of the rotated variance. The findings were discussed in light of a systems approach to evaluation in higher education. Author. (A6, B61, C5, C26)

Michigan State University, Educational Development Program. Student Instructional Rating System manual. East Lansing, Michigan: Office of Evaluation Services, Michigan State University, no date. (a)

The Student Instructional Rating System (SIRS) at Michigan State University is not simply a paper-and-pencil rating instrument but a system for collecting, analyzing, displaying, and interpreting student reaction to classroom instruction and course content. The Student Instructional Rating System Manual (MANUAL) is one of the components of the SIRS system. The MANUAL reported herein describes the units of the system and how each unit can be most effectively used. V. de Wolf. (B4, C20)

Michigan State University, Educational Development Program. Student Instructional Rating System (SIRS) technical bulletin. East Lansing, Michigan: Office of Evaluative Services, University College, Michigan State University, no date. (Duplicated report) (b)

This report described the development, construction, and validation procedures used in the Student Instructional Rating System (SIRS) at Michigan State University. Besides the "general lecture" items, laboratory, recitation, and supplemental items have also been prepared for inclusion with the basic SIRS form. Differences in faculty and student intercorrelation matrices for the general items, for the lab and recitation items, and for supplementary items for specific course content were commented upon in this report. A description of the automated scoring and norming procedures used for SIRS concluded this report. V. de Wolf. (A3, A4, A6, A9, B4, C20, F1, F2, F3, F5, F6, F9, F17, F23, F28, F29, F52, G1, G5)

Miklich, D. R. An experimental validation study of the Purdue Rating Scale for Instruction. Educational and Psychological Measurement, 1969, 29, 963-967.

At the University of Hawaii the author taught two courses: (1) a statistics course which he enjoyed; and (2) an Introductory Psychology for Majors course which included several areas of psychology in which he was relatively unprepared and disinterested. At the end of the semester the Purdue Rating Scale for Instruction (PRSI) was given to both classes. For the first part of the PRSI 2 psychologists predicted the kind of difference, if any, which would be expected between the classes if the students' ratings were valid. It was expected that since the statistics course would be more enjoyable to its students, halo effect would appear in the form of superior ratings by members of that class on scales where no differences were predicted. The results are as follows: the 4 scales where differences were predicted show differences in the predicted direction; 'no difference' predictions were confirmed with 4 scales; and the 2 remaining scales for which no differences were predicted showed tendencies which, while they do not reach customary significance levels, suggested the presence of some halo effect. V. de Wolf. (B61, C43, D28, G3)

Miller, M. T. Instructor attitudes toward, and their use of, student ratings of teachers. Journal of Educational Psychology, 1971, 62, 235-239.

Thirty-six teaching assistants in three freshman courses were assigned to groups on the basis of an attitude scale measure regarding the value they ascribe to student ratings of teachers. Then they were randomly assigned to one of two groups, one of which received feedback from students' ratings during the semester. Differences between attitude and feedback groups were examined using as criterion measures (a) end of semester ratings and (b) final examination scores, adjusted by analysis of covariance using midsemester ratings and examination scores as covariates. Instructors in feedback or attitude groups did not differ significantly on their end-of-semester ratings. In two of the classes there were no significant differences on the final exam scores for the feedback and attitude groups. In the third class, final examination scores were higher for those students whose teachers received feedback from the ratings. Author. (C27, D30, F25, J3, J5)

Miller, R. I. Evaluating faculty performance. San Francisco: Jossey-Bass, 1972.

The publication of this book resulted from a project begun in 1968 by Committee 'C' of the University of Kentucky's chapter of the AAUP. In it are presented different types of faculty appraisal methods. Some of the evaluation methods are as follows: student; classroom visitation; teaching materials and procedures; special incidents; self-evaluation; student appraisal of advising; faculty service and relations appraisal; administrative effectiveness appraisal; performing and visual arts--self-evaluation plus quantified appraisal; professional status and activities appraisal; book appraisal; monograph appraisal; special report appraisal; chapter-in-book appraisal; periodical article appraisal; public service appraisal; and research--self-evaluation plus quantified appraisal. Examples of forms for each of the above mentioned types of evaluations are included. Appendix B includes the results of research on the student evaluation of instruction at the Univ. of Kentucky. V. de Wolf. (A11, B62, D28, F5, F10, F19, F23, F26, F27, F35, F43, F44, F45, F46, F47, F48, F49, F50, G1, G4, G5)

Mueller, R. H., & Roach, P. J. What constitutes an 'ideal' professor? Durham, New Hampshire: Department of Psychology, University of New Hampshire, 1970. (Duplicated report)

The purpose of the present study was to determine what students at the University of Windsor, Canada, considered to be the most and least important characteristics of the 'Ideal' professor, and to compare these results with those obtained by other investigators. Gadzella's questionnaire, "Students' Views of an 'Ideal' Professor," was administered to 225 students at the University of Windsor. The results were almost identical to those which have been reported elsewhere. The high degree of conformity led the authors to believe that perhaps there were no particularly noticeable differences in most and least important criteria across Canadian and American samples. V. de Wolf. (B63, C15, I4)

Mueller, R. H., Roach, P. J., & Malone, J. A. College students' views of the characteristics of an "ideal" professor. Psychology in the Schools, 1971, 8, 161-167.

The purpose of the present study was to determine what students at the University of Windsor, Canada, considered to be the most- and least-important characteristics of the "ideal" professor and to compare these results with those obtained in other studies. Gadzella's 25-item questionnaire, entitled

Students' Views of an "Ideal" Professor, was administered to 642 students. Analyzing the results by Spearman rank order correlation, there existed no significant difference by sex nor no significant difference by student major. These results were very similar to the findings of other researchers with regard to the students' ranking of the most- and least-important criteria. V. de Wolf. (A11, B63, C15, F6, F7, I4)

Murray, H. G. Predicting student ratings of college teaching from peer ratings of personality traits. Paper presented at the annual meeting of the Midwestern Psychological Association, Chicago, May 1973.

Student ratings of teaching were obtained for 36 full-time faculty members in the Department of Psychology, University of Western Ontario, Canada, by means of a 33-item multiple-choice questionnaire. Peer ratings of personality traits were obtained for the same 36 instructors during the 1971-1972 academic year. 11 of the 20 personality traits correlated significantly with mean student ratings of overall teaching effectiveness. A multiple regression analysis, with personality traits as predictor variables and overall teacher rating as the criterion, indicated that a linear combination of 4 personality traits (i.e., leadership, objectivity, extraversion, and anxiety) accounted for 67% of the variance in overall teacher ratings, $\underline{R} = .82$, $\underline{p} < .01$. Finally, correlations were computed between personality traits and mean student ratings for each category of the teacher rating questionnaire. As expected, personality traits were more closely related to questionnaire categories pertaining directly to the instructor (e.g., Communication Skills, Interest Value, Rapport) than to categories pertaining to the textbook, grading system, or course difficulty. V. de Wolf. (B64, D12, H3)

Nelson, C. V., & Cossaart, E. R. The adaptation of the method of summated ratings using normal deviate weights to a form for student evaluation of instructors. Paper presented at the annual meeting of the American Educational Research Association, Chicago, April 1974.

The authors described the development of a student evaluation form at the Department of Mathematical Sciences of Ball State University. Next, they proceeded to relate their method of scoring the evaluation forms using normal deviate weights. The use of normal deviate weights was then compared with integral weights. Even though these 2 correlated well, the authors seemed to prefer the normal deviate system because of its advantage of giving an indication of the position of the response categories on the continuum. V. de Wolf. (A9, A15, B82, C53)

Nichols, M. G. A study of the influences of selected variables involved in student evaluations of teacher effectiveness. Dissertation Abstracts, 1968, 28(8-A), 2908. (Ann Arbor, Michigan: University Microfilms. Order No. 68-1140.)

The purpose of this study was to find whether student, course, or teacher variables affected student ratings of teacher effectiveness. An 18-item questionnaire, which was already used at the State College of Iowa, was employed to rate the professors. Some of the findings were: (1) The student variable of sex produced significant differences in ratings on 10 items of the rating scale, students' class standing affected ratings significantly on 16 items, and 7 items were significantly different in the ratings between students with a 3.00 GPA or above and those with below a 2.00 GPA; (2) The course variable of class size produced significant differences on all 18 items of the questionnaire, classes meeting at different times of the day yielded significant differences on 17 of the items, and the variable of elective or required classes produced significant differences in student ratings of 10 items; and (3)Each of the following--teacher rank, degree, and years of experience at the State College of Iowa--produced significant differences on all 18 items while teacher sex significantly affected ratings on 17 of the items. V. de Wolf. (B83, D1,D4, D9, D27, F2, F5, F6, G1, G2, G5)

Null, E. J., & Nicholson, E. W. Personal variables of students and their perception of university instructors. College Student Journal, 1972, 6(1), 6-9.

This research sought relationships between personal variables--sex, marital status, home background, major field of study, and attitude toward children--of students at Purdue University and their perception of (a) the aggregate behavior of professors and graduate instructors in a team taught course and (b) the behavior of one specific professor from among this group. Students in the sample responded to 3 instruments: the Instructor Behavior Description Questionnaire (IBDQ); a brief biographic inventory; and the Minnesota Teacher Attitude Inventory (MTAI). The MTAI was used in the research because it gave a score indicative of the attitude of a teacher toward the children in his/her classroom. Using single classification analysis of variance, married students gave a significantly higher mean score to the 'one professor' IBDQ score than did unmarried students, and a significant difference existed among students from different major fields for both 'aggregate staff' IBDQ score and the 'one professor' IBDQ score. However, Scheffe tests failed to identify any two means associated with major fields as significantly different from each other. Correlations between both IBDQ scores and the MTAI score were weak. V. de Wolf. (B8, C45, F6, F7, F51, F52, F53)

Null, E. J., & Walter, J. E. Values of students and their ratings of a university professor. College Student Journal, 1972, 6(4, part 1), 46-51.

The study examined relationships between values and selected other variables of students and their ratings of one male university professor. A sample of 192 students enrolled in one large lecture section answered the Allport, Vernon, Lindzey Study of Values, a brief questionnaire, and the first ten items of the Purdue Rating Scale for Instruction. Two-way analyses of variance provided evidence that only a few values are related to ratings by students of one or more dimensions of instructor behavior, but that sex and grade expected in the course are main effects on the evaluation of several of these behavioral characteristics. It was concluded that values of students in general are quite independent of their ratings of a professor. Authors. (B8, C43, E4, F6, F7, F11, F23)

Oles, H. J., & Lencoski, A. Changes in instructor self-ratings resulting from feedback from student evaluations. Paper read at annual meeting of the Southwestern Psychological Association, Dallas, April 1973.

Although student evaluations of teaching performance are now being used on most college campuses, little research has been completed to show their actual effect on the professor. This study tested the hypothesis that a teacher will change his self- evaluation as a consequence of obtaining student evaluative information.

The results confirmed the hypothesis. However, some teachers changed their self-evaluations in a direction exactly opposite of what would have been expected from the student feedback information. Authors. (J2)

Painter, J. J., & Granzin, K. L. Consistency theory as an explanation of students' course evaluation tendencies. The Journal of Experimental Education, 1972, 41(1), 78-81

In an attempt to gain insight into factors influencing college course evaluations, 759 students were questioned at both the beginning and end of the term concerning their feelings about the instructor and course, and about their expectations concerning their grade achievement. Propensity to revise grade expectation over the term did not relate to propensity to revise course evaluation. Those who did revise grade expectation, however, significantly more often than not shifted their evaluation in the same direction. This change in expected grade related more closely to evaluation change than did measures employing grade deserved and actual grade received. Authors. (B36, F3, F23, F24, F26, F27, F32, F33, F36, J6)

Pambookian, H. S. The effect of feedback from students to college instructors on their teaching behavior. Dissertation Abstracts International, 1973, 33(9-A), 4950. (Ann Arbor, Michigan: University Microfilms. Order No. 73-6893.)

In addition to testing the effect of feedback from students to college instructors on the instructors' behavior, this study also attempted to investigate the relationship between change in teaching behavior and prior experience of instructor, sex of instructor, level of student ratings prior to feedback, and the discrepancy between student evaluation of instructor and instructor self-evaluation. The initial level of student evaluation of instruction was found to be a significant source of influence on instructor change after feedback. Instructors evaluated moderately well improved their teaching significantly more in skill, interaction, and rapport when compared to instructors more favorably rated. Instructors whose initial perceptions of their teaching were more favorable than those of their students made more positive change after feedback in the areas of skill, feedback, and rapport than instructors whose initial perceptions of their teaching were less favorable than those of their students. The differences between feedback-receiving inexperienced and experienced instructors were not significant. But when students described their instructors, results indicated that female instructors showed significantly positive change on the dimensions of skill and interaction in their teaching when compared to male instructors. V. de Wolf. (B19, C8, D1, D2, H1, J3)

Pambookian, H. S. Initial level of student evaluation of instruction as a source of influence on instructor change after feedback. Journal of Educational Psychology, 1974, 66, 52-56.

In early October 1971, 252 students in 13 introductory and educational psychology sections responded to a Student Opinion Questionnaire containing measures of seven stable dimensions on college teaching. Ten days later, the instructors, who had been grouped according to the initial level of student evaluation, received feedback. In December, 231 students responded again to the Student Opinion Questionnaire. The students' initial evaluation of instruction was found to be a significant source of influence on instructor change after feedback. Instructors who were evaluated moderately well improved their teaching significantly on skill, interaction and rapport, when compared to instructors rated more favorably. They showed trends toward decreasing work overload and improving rapport when compared to instructors rated more unfavorably. Author. (B19, C8, J3)

Perkins, E. R. Relationships among empathy, genuineness, nonpossessive warmth, and college teacher effectiveness and selected characteristics. Dissertation Abstracts International, 1972, 32(9-B), 5454-5455. (Ann Arbor, Michigan: University Microfilms. Order No. 72-9405.)

This study sought to determine if a significant relationship existed among the specified variables of empathy, unconditional positive regard and the self-congruence of selected college teachers and their teacher effectiveness as adjudged by students. Two of the results were: (1) If college professors scored high on the facilitative conditions of empathy, genuineness, and nonpossessive warmth, then they tended to receive higher ratings of effectiveness as adjudged by their students; and (2) Grades increased with the greater degree of empathy, genuineness, and nonpossessive warmth which was present. Both of these were confirmed at the .001 level of significance. V. de Wolf. (B84, C54, D1, D3, D4, D12, D14, D27, D36, H4)

Permut, S. E. Cue utilization patterns in student-faculty evaluation. The Journal of Psychology, 1973, 83, 41-48

Judgment by students or faculty effectiveness have relied, for the most part, on subjectively expressed responses to a series of "effectiveness traits" (cues). Simple averaging schemes have summarized the aggregate judgments, often resulting in (a) treating all cues as if they were of equal importance, and (b) inattention to the effects or measurement of individual differences of judges.

Multiple regression was used in this study to derive (empirically) individual cue utilization patterns (i.e., weights) for 10 effectiveness traits. Ss were compared to their subjectively expressed judgment scheme. Results indicate that Ss were moderately successful in expressing their actual cue utilization patterns; however, different judgment schemes were clearly observed. Author. (A15, B3, I4)

Perry, R. R. Evaluation of teaching behavior seeks to measure effectiveness. College and University Business, 1969, 47(4), 18 and 22. (ERIC: EJ009135)

Reported in this article was a University of Toledo's study on criteria of effective teaching. Faculty, students, and alumni/ae were contacted and requested to identify effective teaching behaviors. A 60-item survey resulted. Faculty, students, and alumni/ae then determined the relative importance of the items. The final ranking by all 3 groups was included in this report, as was a value factor for each item (this value factor accounted for the demonstrated differences in importance of each criterion computed from weighted raw score totals). V. de Wolf. (A9, A15, B66)

Perry R. R., & Baumann, R. R. Criteria for the evaluation of college teaching: Their reliability and validity at the University of Toledo. In A. L. Sockloff (Ed.), Proceedings: The first invitational conference on faculty effectiveness as evaluated by students. Philadelphia, Pa.: Measurement and Research Center, Temple University, 1973. Pp. 183-212.

This paper contained two complementary sections. The first section, written by R. R. Perry, described the development of an instructional evaluation instrument at the University of Toledo. He reported the results of ranking the criteria established at the University of Toledo with those obtained at 3 other public universities and concluded that there was no significant disagreement, among the reporting categories selected for study, about the rank order importance of these criteria. In the second paper, R. R. Baumann prepared a rating scale, entitled the Student Perception of Teaching Effectiveness Scale, on the basis of the judgments of teaching effectiveness as described in the companion paper. In an attempt to diminish the bias present by size of class and instructional level, 13 norm groups were established. The rating of each course then was made relative to the ratings of other courses of the same size and class level. In general, smaller classes gave better ratings. Also, the higher the instructional level, the better the ratings. Relationships between the scores from the scale and certain variables, i.e., student sex, required/elective nature of the course, reported GPA, and expected grade, were then explored. The first two variables, sex and required/elective, could be diminished through norming procedures. The variables of GPA and expected grade seemed to have considerable relationship with ratings. V. de Wolf. (A6, A9, A12, A13, A15, B29, B30, B31, B66, C28, F2, F5, F6, F23, G1, G5, H5, K3)

Pfeiffer, M. G., & Rosbach, L. A. Teaching performance criterion development through scaling: Teacher-student points-of-view analysis. Perceptual and Motor Skills, 1969, 28, 755-766.

30 operational tasks of psychology professors were scaled by the 14 students and 10 of the 14 faculty of the same department using an adaptation of the pair comparisons method. The stimuli, in the form of operational statements of classroom teacher activities, were compared two at a time by Ss who made judgments in terms of the amount of similarity between activities. These judgments were then transformed into correlation coefficients and factor analysed. Factorial congruence between faculty and student judgments was demonstrated. The classroom activities of the psychology professors at the urban college may be described in terms of 8 basic dimensions: "Knowledge Dissemination," "Teacher Dynamism," "Classroom Administration," "Control of Student Behavior," "Environmental Regulation," "Advisory Guidance," "Information Dissemination," and "Teacher-Student Feedback." Authors. (A3, A6, A9, C30, H3)

Pfeiffer, M. G., Lehmann, W., & Scheidt, U. Cross-cultural scaling studies in the development of probabilistic teaching performance criteria anchored to utility and time scales. Technical Reports Number 5. (Rev. ed.) Philadelphia, Pennsylvania: Psychological Laboratory, La Salle College, 1970.

This report is in two parts.

(1) Chapter II reports the expected amount of teaching time wasted in the classroom and compares an American and a German psychology department. Estimations of teaching performance effectiveness were given by German and American psychology students on each of eight psychology teachers of their departments for each of eight classroom teaching performance factors. These values were converted to probabilities of ineffective teaching performance and multiplied with teaching time needed for each classroom teaching performance factor. By this procedure the teaching performance measurement of expected amount of time wasted in the classroom was developed.

(2) Chapter III reports an investigation into the subjective expected utility of teaching performance. This investigation was also cross-cultural in that it compared American and German psychology departments.

Performance effectiveness estimations were given by German and American psychology students on each of eight psychology teachers of their departments for each of eight classroom teaching performance factors. These values were converted to probabilities and multiplicatively combined with the utility of teaching performance estimates of students to yield the subjective expected utility of classroom teaching performance. Intra- and inter-cultural similarities and differences are discussed. Authors. (A11, A16, C30, D15, F39)

Pfeiffer, M. G., Lehmann, W., & Scheidt, U. Methodological investigation of perceived structure of college teaching across two cultures. Perceptual and Motor Skills, 1972, 35, 619-626.

4 multidimensional scaling analyses based on similarity ratings among 30 operational statements of college-level psychology teaching activities (stimuli) using samples of psychology students and teachers from the USA and Germany as raters resulted in 7 teaching factors common in varying degrees to both cultures and one in each culture with no counterpart in the other. The paired German-American factors ordered from most to least agreement in accordance with the proportion of their highly loaded stimuli held in common by corresponding factors across cultures were: Teacher (Lecturing) Dynamism, Dynamik des Vortragens; Classroom Administration, Durchsetzung der Ordnung; Information Dissemination, Ausgabe von Informationen; Knowledge Dissemination, Verarbeitung von Wissen; Advisory Guidance, Studentische Beratung; Environmental Regulation, Regulierung von Randbedingungen; Teacher-Student Feedback, Lehrer-Student Rueckmeldung; Control of Student Behavior, Verwendung von audiovisuellen Hilfsmitteln. Analyses of the degree of student-teacher agreement of factors within cultures resulted in a greater degree of within culture agreement than the degree of cross-cultural agreement shown among the 8 teaching factors. The clarity of the current results may be interpreted as indicating that the method employed possesses potential for establishing the degree of perceptual generality and meaningfulness of teaching performance criteria. Authors. (A9, A16, C30, H3)

Plant, W. T., & Sawrey, J. M. Student ratings of psychology professors as teachers and the research involvement of the professors rated. The Clinical Psychologist, 1970, 23(4), 15-16 and 19.

Results of an eight-item student ratings scale were correlated with five different research involvement scores for 32 professors. Of 27 correlation coefficients reported for the nine teaching effectiveness scores (8-items plus total mean) and the first 3 research involvement scores (publications, papers read, and theses chaired), 12 are negative and 15 positive, with 5 significantly different from zero beyond the .05 level. None of the significant coefficients were very large. Mean teaching effectiveness ratings for those on versus those not on research grants were also computed. Of the 18 t ratios reported, only 4 were significant beyond the .05 level. The authors concluded that the present study provided no systematic empirical support for either contention regarding the relationship between teaching effectiveness and research involvement. V. de Wolf. (A9, C29, D6, D7, D31, D32)

Price, J. R., & Magoon, A. J. Predictors of college student ratings of instructors. Paper presented at the annual meeting of the American Psychological Association, Washington, D. C., September 1971.

A large sample of college students completed a thirty-five item course evaluation instrument. The instrument could be divided into two sections: (1) eleven course and student characteristics which presupposed no evaluative judgments; and (2) twenty-four rating items bearing on the course and instructor, all evaluative judgments. The results of a canonical analysis revealed four important correlational relationships. These suggested that students rate courses on the basis of instructor impact, workload, course structure, and whether the course is an elective or not. Important predictors were the expected grade in a course and availability of the instructor outside class (all members of this sample sought such help). Authors. (B15, C43, D33, D34, F5, F6, F14, F21, F23, F55, F56, F57, G7)

Pritchard, W. M. Student evaluation of college physics teaching. Journal of Research in Science Teaching, 1972, 9(4), 383-384.

A university-wide program of student evaluations of faculty teaching was initiated by Old Dominion University in the fall semester on the 1970-71 academic year. Data available from the administration of the questionnaire to 34 physics courses in the 1970-71 academic year were analyzed. It was reported that higher level courses had higher ratings, with averages for 300, 400, and 500 level courses being essentially the same. Since physics faculty members normally taught both 100 level courses and advanced courses, the observed correlation between student ratings and course level could not be explained as due to the assignment of poorer teachers to introductory courses. V. de Wolf. (B5, G4)

Pugh, R. C., & Richmond, M. G. The consistency of student ratings of instruction. Paper read at the annual convention of the National Council for Measurement in Education, Chicago, April 1974.

The consistency of student ratings of instruction within a class was assessed by cluster analyzing students based on their responses to a 38-item instrument. Intraclass correlation coefficients were used to compare the consistency of ratings within a class to the consistency of ratings among classes. For this study, 32 classes were randomly selected in an arts and sciences college during the first semester, 1972-73. The median number of student clusters within a class was three. The number of clusters within a class ranged from five to one. Only one of the 32 classes had a single cluster. Intraclass correlations were moderately low. Recent tabular summaries which have reported item means for a class do not reflect this internal inconsistency. Authors. (A11, A17, C56)

Quereshi, M. Y. Letter on "Teaching effectiveness and government awards." Science, 1968, 161, 1160.

According to Quereshi, there were several deficiencies in the data presented in Bresler's article which seemed to dilute the force of Bresler's argument. Some of these deficiencies were in the method of collecting data--or at least in those methods as presented in the article--and others were in the analysis of the data. V. de Wolf.

Quereshi, M. Y., & Widlak, F. W. Students' perception of a college teacher as a function of their sex and achievement level. The Journal of Experimental Education, 1973, 41(3), 53-57.

Sixty-seven male and seventy-three female undergraduates rated self, father, mother, and a male teacher on MARS. They also rated the teacher on a 17-item scale (RST) at the end of the course and took five teacher-made achievement tests during the semester. Students were divided into high and low achievement levels (AL), using the median as the cutoff. Two-way multivariate analysis of variance showed significant ($p < .0001$) differences in teacher ratings on MARS and RST factors between the sexes and between high and low AL Ss ($p < .05$). Also, significant sex x AL interaction ($p < .05$) was found for the RST data. Authors. (A6, A11, B68, C32, C33, D12, F6, F12, F16, J5 J6)

Rayder, N. F. College student ratings of instructors. The Journal of Experimental Education, 1968, 37(2), 76-81. (ERIC: ED021527)

In a study to examine the nature of student ratings,over five thousand students rated their instructors on each of three criteria. The ratings were then correlated with selected student characteristics and instructor characteristics. The technique used was multiple linear regression. Generally, mean student ratings of instructors were more related to instructor characteristics than to such student characteristics as age, grade level, GPA, course grades, sex, or grade(s) previously received from instructor. Older, more educated and more experienced instructors were rated lower than less experienced instructors. V. de Wolf. (B7, C34, D1, D2, D3, D4, D9, D10, F1, F2, F4, F5, F6, F7, K2)

Rees, R. D. Dimensions of students' points of view in rating college teachers. Journal of Educational Psychology, 1969, 60, 476-482.

A factorial study was designed to yield "points of view" or "idealized individuals" with respect to the rating of college teachers. The 11 types of teachers used as stimuli were selected as representative of seven academic areas and were rated by 65 student Ss on 20 semantic differential scales. An obverse factor analysis yielded 20 bipolar factors, eight of which, after orthogonal Varimax rotation, were selected as significant and were then correlated with a number of outside variables to assist in their interpretation. Seven of the eight factors were identified. The identified factors, which represent different points of view in rating college teachers, were labeled "socioeconomic," "racial," "social studies aptitude," "class in school," "masculine sophistication," "social disposition," and "emotional instability" factors. Author. (A6, B69, C24, D10, F1, F2, F6, F7, F12, F13, F30, F52, F58, F59, F60, F61)

Roberts, D. Y. Research paper. Gainsville, Florida: Institute of Higher Education, University of Florida, 1971. Cited by O. S. Lefforge, Inservice training as an instrument for change. Gainsville, Florida: Institute of Higher Education, University of Florida, 1971. Pp. 39-45. (Duplicated report) (ERIC: ED055577)

The research paper cited by Lefforge presented the five most dominant characteristics of the best college instructors and the five most dominant of the worst college instructors as designated by students. These designations were shown to cluster around 7 headings. A percentage breakdown was given in Lefforge's article. V. de Wolf. (A10)

Rodin, M., & Rodin, B. Student evaluations of teachers. Science, 1972, 177, 1164-1166. (ERIC: EJ064736)

Teaching assistants in an undergraduate calculus course were rated by their students. There were 12 sections, two of which were taught by the same instructor. Besides the students' evaluations of their instructors, two additional measures were obtained for each of the sections: initial ability in calculus and amount learned by the students (an objective measure of teacher effectiveness). The partial correlation between objective and subjective measures of teaching ability, with initial ability held constant, was equal to -.746. The ordinary correlation between the two measures was r = -.754. The explanation for the negative correlation between the amount learned from an instructor and the students' evaluation of his teaching performance was not obvious. After positing several possible explanations, the authors concluded that if how much students learn was considered to be a major component of good teaching, it must be concluded that good teaching was not validly measured by student evaluations in their current form. V. de Wolf. (F3, F25)

Rosenshine, B., Cohen, A., & Furst, N. Correlates of student preference ratings. Journal of College Student Personnel, 1973, 14, 269-272.

The purpose of this study was to analyze the data from the 38-item Course Evaluation Survey in order to determine whether certain demographic variables were related to the student ratings and which rating variables were related to certain preference criteria. The correlations between the demographic variables and each of the preference criteria yielded two results of interest: first, those who expected a higher grade in the course tended to give more favorable ratings on the preference criteria (r = .09 to .27); and, second, there was a tendency for more experienced students (those who were older, had more courses in the field, or were upperclassmen) to give higher ratings, although only one of those correlations was as high as .10. The lack of any meaningful correlations between the demographic variables used in this study and the 5 preference criteria cross-validated results obtained in other studies of this type. Also included in this article were the intercorrelations among the 5 preference criteria and the correlations between the preference criteria and certain items of the Course Evaluation Survey. V. de Wolf. (B70, C35, F1, F2, F5, F6, F9, F23, F52)

Rous, S. N., Bamford, J. C., Jr., Gromisch, D., Rich, H., Rubin, S., & Sall, S. The improvement of faculty teaching through evaluation: A follow-up report. Journal of Surgical Research, 1973, 13, 262-266.

A follow-up on a program to improve the quality of faculty teaching through evaluation was presented. Sixteen faculty members on the Clinical Faculty of a medical school who participated in an evaluation learned the results of that evaluation and then were re-evaluated. Statistically significant improvement was demonstrated in their overall teaching profiles and in 4 of the 5 separate teaching categories. Authors. (J3)

Rous, S. N., Bamford, J. C., Jr., Gromisch, D., Rubin, S., & Sall, S. The improvement of faculty teaching through evaluation: A preliminary report. Journal of Surgical Research, 1971, 11, 311-315.

The authors discussed the development of their evaluation questionnaire for use by students in a medical school. V. de Wolf. (A9)

Rubinstein, J., & Mitchell, H. Feeling free, student involvement, and appreciation. Proceedings, 78th Annual Convention, APA, 1970, 623-624.

The primary concern of this study was to investigate the relationships among feeling of control, exercise of control, appreciation of instructor, and appreciation of the course. Grades on departmental exams and grades on instructor-assigned tasks were examined in relationship to all of the aforementioned variables. A correlation analysis by instructor showed that the feeling of control index was highly related to appreciation of the instructor, and also related to exercise of control and course appreciation. However, exercise of control was not related to any of the other variables examined. This suggested that feeling of control, whether the control was exercised or not, was an important factor in course and instructor evaluation. Although grades were related to course appreciation, they were not related to appreciation of the instructor or to the control indexes. V. de Wolf. (A6, F3, F14, F25, F45, G6)

Scott, C. S., & Thorne, G. Assessing faculty performance: A partially annotated bibliography. Monmouth, Oregon: Faculty Assessment Project, Teaching Research Division, Oregon State System of Higher Education, 1974. (Duplicated report)

139 articles, papers, and books written over the past 20 years were included in this report. Approximately 50% were abstracted. Topics covered were wide-ranging and included all levels of education from public school staff to university professors. Many of the varied aspects of assessment were touched upon: theories and potentially useful models; actual research studies; cost of effectiveness analysis; a critique of the American Council on Education's institutional ratings; ratings of the ideal teacher; institutional policies in faculty evaluation; faculty evaluation by peers, administrators, and students; and faculty load evaluations. V. de Wolf. (A2)

Scott, O., Halpin, G., & Schnittjer, C. Student characteristics associated with student perceptions of college instruction. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, April 1974.

The purpose of the study was to ascertain relationships between selected student characteristics and student perceptions of 6 aspects of college instruction--relevance of instructional objectives, quality of classroom interpersonal relations, use of instructional resources, student motivation, meaningfulness of subject matter content, and procedures for measuring and evaluating attainment of instructional objectives. The student characteristics selected were sex, creativity, life history, and academic achievement status. Some selected results were: (1) Student sex differentiated the group in student motivation and content meaningfulness; (2) No differences were found between highly creative and less creative students on the 6 aspects; (3) There was a lack of relationship between life history and perceptions of instruction; and (4) For male students there was a relationship between academic achievement and perception of human relations-values (for this analysis perception of classroom interpersonal relations was subdivided into human relations-values and human relations-student participation in decision making). V. de Wolf. (A8, C57, F3, F5, F6, F12, F72)

Sharon, A. T. The effect of instructional conditions in producing leniency on two types of rating scales. Dissertation Abstracts, 1969, 29(8-B), 3124. (Ann Arbor, Michigan: University Microfilms. Order No. 69-2520.)

See Sharon and Bartlett (1969) for an abstract of this dissertation. V. de Wolf.

Sharon, A. T. Eliminating bias from student ratings of college instructors. Journal of Applied Psychology, 1970, 54, 278-281.

A forced-choice scale of teaching effectiveness was constructed by obtaining preference and discrimination indexes on 300 descriptive statements of teacher behavior. The two indexes were used to group 60 of the statements into 15 forced choice tetrads. The rating scale was used by 504 introductory psychology students to rate 14 graduate teaching assistants under four different instructional conditions while another group of 542 students rated the same teaching assistants with a conventional graphic scale. It was found that the instructional conditions had no effect on the forced-choice ratings but had a significant effect on the graphic ratings. It was concluded that the forced choice scale is resistant to bias occurring in student ratings of college instructors. Author. (A3, A5, A7, B71, F20)

Sharon, A. T., & Bartlett, C. J. Effect of instructional conditions in producing leniency on two types of rating scales. Personnel Psychology, 1969, 22, 251-263.

The objectives of this study were to determine the effects of several instructional rating conditions on leniency on a graphic and forced-choice rating scale and to test Wherry's theory of rating with respect to the leniency bias. College students rated their instructors under instructional conditions indicating that the ratings (1) will be anonymous and will be used for research purposes only; (2) may be used for evaluation purposes; (3) will be identified by having the rater place his name on the rating form; or (4) will have to be explained to the ratee by the rater. A significant leniency effect was found with the graphic ratings which were to be used for evaluation purposes and those that had to be justified to the ratee. It was concluded that the forced-choice scale was quite resistant to leniency bias and that Wherry's theory has been supported with the results of the graphic ratings. Authors. (A3, A5, A7, A11, B71, F20)

Slobin, D. Y., & Nichols, D. G. Student rating of teaching. Improving College and University Teaching, 1969, 27, 244-248.

Student rating of teaching commonly has met opposition and hostility from faculty and yet after a few years experience most faculties have come to accept it and even praise it. In this article the authors voiced certain typical objections to student evaluation of teaching and then refuted each objection with a review of the results of past studies in the area of student ratings of instruction. V. de Wolf. (A2, A6, B1, B8, B19, B53, B54, D1, F1, F2, F3, F5, F6, F8, F23, G1, G5)

Smithers, A. Some factors in lecturing. Educational Review, 1970, 22, 141-150. (a)

Four hundred and thirty-one students in their second year at the University of Bradford rated 50 possible characteristics of the ideal lecturer. Responses were inter-correlated and factor analysed by the varimax procedure. Twelve factors isolated in this way were interpreted as questions which students ask themselves in evaluating lecturers. Students in all fields of study were agreed that the ideal lecturer is an authority in his subject and can expound it clearly; that he thoroughly prepares his lectures, gives them an obvious framework and is ready to respond to questions. In other respects there were differences in emphasis according to field of study: students of applied science and engineering appeared to look towards lecturers for information, social scientists, for stimulation. The differences in emphasis suggest that teaching and lecturing abilities may be more specific than is commonly supposed. This must be recognised in any training programmes devised for new lecturers. Author. (A6, B72, F7, I2, I4)

Smithers, A. What do students expect of lecturers? Universities Quarterly, 1970, 24, 230-236. (b)

A 50 item lecturer characteristics questionnaire rating the ideal lecturer, the Eysenck Personality Inventory, and Rokeach's dogmatism scale were administered to a group of students at the University of Bradford, England. It was hypothesized that students of different personality types would have differing concepts of the ideal lecturer. The author concluded his analysis of the data by showing that students' conceptions of the ideal lecturer did vary with personality types, but these differences were comparatively few and comparatively slight. On the whole, students of different personality types tended to hold rather similar standards for lecturer behavior. V. de Wolf. (B72, F12, I3, I4)

Sockloff, A. L. Instruments for student evaluation of faculty: Ideal and actual. In A. L. Sockloff (Ed.), Proceedings: The first invitational conference on faculty effectiveness as evaluated by students. Philadelphia, Pa.: Measurement and Research Center, Temple University, 1973. Pp 132-151. (a)

The bulk of this article discussed the ingredients necessary for and the construction of faculty evaluation instruments. At the end the author reported the results of his survey of American institutions of higher education. This survey requested information about uses, financial support, sources of items, and methods, as well as copies of instruments, exemplary individual summary sheets, and technical reports describing the construction of instruments used for faculty evaluation by students. The 900 responses yielded about 500 instruments. After discussing his observations on the materials he received, the author concluded that, "perhaps, if the instruments had been researched early in their development, a good 99% would not be around today." V. de Wolf. (A2)

Sockloff, A. L. (Ed.) Proceedings: The first invitational conference on faculty effectiveness as evaluated by students. Philadelphia, Pa.: Measurement and Research Center, Temple University, 1973. (b)

This book contained the twelve papers read at the First Invitational Conference on Faculty Effectiveness as Evaluated by Students. Papers included in this annotated bibliography were: Centra, 1973e; Aleamoni, 1973b; Sockloff, 1973a; Hoyt, 1973a; Warrington, 1973; Perry & Baumann, 1973; McKeachie, 1973; and Clark & Blackburn, 1973. V. de Wolf.

Sorey, K. E. A study of the distinguishing personality characteristics of college faculty who are superior in regard to the teaching function. Dissertation Abstracts, 1968, 28(12-A), 4816. (Ann Arbor, Michigan: University Microfilms. Order No. 68-8502)

50 college instructors were rated on the Purdue Rating Scale for Instruction and divided into two groups--the upper 26% and the lower 26%. Three hypotheses were tested: (1) Superior teachers, as rated by their students, would score in the socially valued direction on a greater number of the Guilford-Zimmerman traits than would inferior instructors; (2) Superior professors would show a more positive self-concept, that is, rate themselves more positively on the Self-Rating Scale (a scale designed to allow self-assessment on the same traits as the Guilford-Zimmerman), than would the inferior teachers; and (3) Superior teachers would show more accuracy in their self-rating, as measured by the discrepancy between self-rating and Guilford-Zimmerman scores, than would inferior instructors. Hypotheses 1 and 3 were not confirmed and only provisional support was obtained for hypothesis 2. V. de Wolf. (A10, C43, D12, H1)

Spencer, R. E., & Aleamoni, L. M. A student course evaluation questionnaire. Journal of Educational Measurement, 1970, 7, 209-210.

The Illinois Course Evaluation Questionnaire, based upon 50 items and used at 13 different institutions, was described in this article. V. de Wolf. (A9, A11, B11, B18, B19, B37, B48, B70, B73, B74, B75, B76, B77, B78, B79, C5)

Stallings, W. M., & Singhal, S. Some observations on the relationships between research productivity and student evaluations of courses and teaching. The American Sociologist, 1970, 5, 141-143. (ERIC: ED027838)

Given the somewhat contradictory and inconclusive findings of previous studies, it was the purpose of this study to investigate further the relationships between teaching effectiveness and research productivity (as defined by scores derived from a weighted combination of the number of books, articles, technical reports and bulletins, and book reviews published within a given time period). With the data from 2 institutions, student ratings of courses and instructors were correlated with the research productivity index (RPI). By using a highly reliable instrument for student evaluation, the University of Illinois ratings had small, statistically significant, linear correlations with the RPI. However, with an instrument of unknown reliability, the Indiana University data showed no statistically significant correlations between RPI and either global or comparative measures of teaching effectiveness. At both institutions there was a significant relationship between academic rank and RPI. V. de Wolf. (B3, B9, C5, C36, D4, D6, D8, D10, D11, D27, D32, G1, G4)

Swanson, R. A., & Sisson, D. J. The development, evaluation, and utilization of a departmental faculty appraisal system. Journal of Industrial Teacher Education, 1971, 9(1), 64-79. (ERIC: EJ048740)

It was the purpose of this study to (1) develop a theoretically sound faculty appraisal system (FAS), (2) select or develop appropriate instruments, and (3) to develop a scoring system. Student ratings, peer ratings, and chairman/woman ratings were the 3 parts of the FAS. Rank-difference correlations between these 3 types of ratings ranged from .06 to .88 for the 2 quarters of the study. The theoretical elements of the FAS and proposed independence of the 3 adopted measures were partially supported by these results. Utilization of these results for promotion, tenure, and salary decisions was then discussed. V. de Wolf. (A9, B11, C5, H3, H8)

Taylor, R. E. An investigation of the relationship between psychological types in the college classroom and the student perception of the teacher and preferred teaching practices. Dissertation Abstracts, 1969, 29(8-A), 2575-2576. (Ann Arbor, Michigan: University Microfilms. Order No. 69-2235.)

The purpose of this study was to ascertain the relationship between psychological types as described by Jung (as measured by the Myers-Briggs Type Indicator [MBTI]), and (1) the way in which the student values his/her teacher and (2) the student's preference for student-centered or teacher-centered instruction (as measured by the Classroom Preference Test [CPT]). To represent the difference between the teacher and student psychological types as a single score, the Euclidian distance between the two MBTI profiles was computed. When the difference scores between the student and teacher psychological types were correlated with the scores from the Classroom Observation Record (an instrument which measured the difference between the student rating of the Ideal Teacher and the Actual Teacher), the relationship was found not to be significant. When the set of 4 MBTI index scores for each student were correlated with his/her 4 factor scores and the Cumulative Factor Score of the CPT, the relationship was observed to be significant in 4 of the 6 classrooms in the study's sample. V. de Wolf. (A6, A11, B71, D12, F12, F41, I1)

Touq, M. S., & Feldhusen, J. F. The relationship between students ratings of instructors and their participation in classroom discussion. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, February 1973.

The general hypothesis of this research is that student participation in classroom discussion is rewarding and that it reinforces favorable attitudes toward the instructor. A total of 480 undergraduates rated their instructors. Instructors identified high and low participants, and instructors were rated as high and low facilitators of discussion by expert observers. No difference between ratings of high and low participant students were found, but instructors who were rated as high facilitators by experts were also rated higher by students. Authors. (C43, D16, E6, F45, H7)

Touq, M. S., Feldhusen, J. F., & Halstead, J. Criterion-referenced validity of student ratings of instructors. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, February 1973.

The purpose of this research was to assess the criterion-referenced validity of student ratings of instructors. A total of 480 undergraduates rated their instructors using a special rating scale designed to parallel the Flanders Interaction Analysis Categories. Expert observers also rated the instructors using the standard form of the Flanders Categories. Mean student ratings for instructors were correlated with expert observers' scores. Significant correlations were found between ratings for four categories. These results were interpreted as revealing some criterion-referenced validity for student ratings. Authors. (F45, H7)

Treffinger, D. J., & Feldhusen, J. F. Predicting students' ratings of instruction. Proceedings, 78th Annual Convention, APA, 1970, 621-622.

This study examined the correlates of students' ratings. Several different kinds of student variables were used, including cognitive and personality characteristics, specific and general measures, and course performance indexes. The Purdue Rating Scale for Instruction (PRSI) was administered twice, once during the first and once during the last week of the course. The final PRSI ratings were categorized using 3 classes of items derived for the initial, generalized ratings: instructor; objectives and testing measurement; and instructional methods. A summary or overall rating score was also obtained, combining the 3 item classes. Each of these 4 item classes was then used in turn as the dependent variable in a stepwise multiple-regression analysis in which the 20 cognitive, personality, attitude, and academic variables were potential independent variables. The results indicated the relative importance of the generalized precourse ratings (and particularly, generalized ratings of instructional methods), in comparison with the other student variables, as predictors of specific course and instructor ratings. V. de Wolf. (B8, C43, F3, F12, F16, F53, F63, F64, K2)

Turner, R. L., Evans, J. H., Hale, T. A., Cairns, S. G., & Maleski, F. D. How do student characteristics affect their evaluations of instructors. Education Bulletin, School of Education, Indiana University, 1969, 45, 47-97.

The following series of studies portray student ratings of college instructors as the result of interactions between student and instructor characteristics. In order to examine this interaction, characteristics of most and least preferred instructors were identified by students who possess high and low achievement anxiety, and were high and low in academic success. Whereas the more anxious student appears to place greater emphasis on personal-social instructor behavior, the less anxious student values task-oriented, and intellectual instructor characteristics. By comparison, failing students value the controlled, relaxed, and accepting instructor, while the high scholarship student prefers a stimulating, clear, and open-minded instructor. In addition to describing methods by which an instructor may gather and use meaningful student ratings, pertinent implications for changing instructor behavior in light of certain student characteristics are reviewed. Authors. (A10, C58, F2, F12, F16)

University of Washington, Evaluative and Counseling Service. An overview of the teaching evaluation at the University of Washington. Seattle: Evaluative and Counseling Services, University of Washington, 1973. (Unpublished paper)

The history and administrations of the University of Washington's student ratings of teaching effectiveness program was reviewed. V. de Wolf. (A2, B1)

Usher, R., & Hanke, J. The "Third Force" in psychology and college teacher effectiveness research at the University of Northern Colorado. Colorado Journal of Educational Research, 1971, 10(2), 2-9.

The authors summarized the results of 5 studies, completed in the 2 year period from 1969-1971, which explored Third Force hypotheses as they related specifically to effective college teaching in the college classroom situation. The topics covered by the studies were: (1) The relationships between judged effectiveness as a college teacher and the instructors' inferred conceptual systems orientation and perceptions regarding themselves, other people, and the task of teaching; (2) The relationships between judged effectiveness and 4 inferred perceptual dimensions of the healthy personality; (3) The relationships between judged effectiveness and tactual body accessibility and verbal self disclosure; (4) In the fourth study cited, an investigation was made of college teaching effectiveness by compiling the findings used in the first 3 studies; and (5) The last study investigated the relationship between each student and his/her teacher on a one-to-one basis. V. de Wolf. (A2, B7, D1, D3, D10, D12, D19, D23, D24, D25, D26, D37, D38, F2, F12, F19, G1, G5, H1, H2, H3, H6)

Van Horn, C. An analysis of the 1968 Course and Instructor Evaluation Report. Institutional Research Bulletin #2-68. Lafayette, Indiana: Measurement and Research Center, Purdue University, no date.

This analysis of the 1968 Course and Instructor Evaluation Report, a 14-item rating scale, was shown to measure 5 aspects of students' opinions. Most of the variance in students' responses was accounted for in a single general factor with which all of the questions were correlated, and which constituted a one-dimensional "good-bad" evaluation. Differences in ratings attributable to the rank of the instructor were present in some of the questions but to an extent so slight that separate norms were not justified. Differences between departments has not been investigated and evidence suggesting that such investigation would be useful was not seen. There was evidence, however, that class size was systematically related to the responses. The reliability of the rating scales, regarded as a single additive measure, was high. There was evidence from two sources that most of the individual items elicited stable responses. Finally, norms for interpreting median values published in the Course and Instructor Evaluation Report were presented. V. de Wolf. (A6, A11, A13, C46, D4, G1)

Very, P. S. Real and ideal characteristics of the teacher-student relationship. Perceptual and Motor Skills, 1968, 27, 880-882.

A teacher-student questionnaire was administered to a group of 144 basic psychology students. Correlations between the ideal and real sorts indicated the following results. On one hand, students felt that improvements could be made by teachers in the areas of: inspiring students, allowing greater freedom of both teacher and student opinion in class, being more "available" to the students, being more affected by the background and interests of the students in their courses, and providing more than just "book knowledge." On the other hand, students felt that teachers were justly strict with respect to deadlines, agreed that more student participation in class was needed, admitted that students tended to take unfair advantage of permissive professors, and agreed that most students unfortunately tended to get by the easiest way possible. V. de Wolf. (A9, I1)

Villano, M. W., Rosenstock, E. H., & Estes, C. A decade with a student course evaluation form at a major university. Paper presented at the annual meeting of the American Educational Research Association, Chicago, April 1974.

It was the purpose of this present paper to report an investigation of the current reliability and internal validity of Pennsylvania State University's Course Attitude Questionnaire (CAQ), the temporal stability of its originally-claimed factor structure, and its correlation with some important educational variables. The authors reported that there was a discernable inflationary trend in CAQ ratings over the past decade. General increases in the CAQ subscale and total scores were noted for the 10 year time interval covered. An inspection of the cumulative norms through 1973 revealed there was not a sufficient spread in the class means for the scores to discriminate adequately except at the extreme points (10th and 90th percentiles) of the distribution. Multivariate analysis of variance of the CAQ subscale scores for 206 class-sized groups indicated significant main effects for class size, course level, and instructor rank with significant size by level and level by rank interaction effects. It was concluded that the factorial invariance of the CAQ has not been demonstrated and that a strong relationship existed between the CAQ subscale and certain classroom variables--a relationship that ought to be minimal or zero in a highly valid instrument. V. de Wolf. (A2, A6, A11, A13, B10, C59, D4, G1, G4)

Walker, B. D. An investigation of selected variables relative to the manner in which a population of junior college students evaluate their teachers. Dissertation Abstracts, 1969, 29(9-B), 3474. (Ann Arbor, Michigan: University Microfilms. Order No. 69-786.)

The purpose of this study was to investigate some selected variables which might be relative to the manner in which community junior college students evaluate the effectiveness of their teachers. Some of the conclusions reached in this study were: (1) Students tended to rate teachers in the direction of their stated anticipated grade; (2) Teachers of courses which students stated were similar to what they had expected and, also, teachers of courses which the students believed to be fulfilling their needs tended to receive higher ratings; (3) There was no relationship between the level of course difficulty as perceived by the student, sex of student, student's class, or sex of teacher and student ratings of teaching effectiveness; (4) Older students tended to rate teachers higher than did younger students; (5) More experienced teachers tended to get higher ratings than did less experienced teachers; (6) Female students rated female teachers significantly higher than they rated male teachers; and (7) Teachers of mathematics and science tended to get higher ratings than teachers of other subjects. V. de Wolf. (B85, C43, D1, D2, D10, F1, F2, F6, F23, F26, F73, F74)

Walsh, G. V. One in five made us think. Improving College and University Teaching, 1972, 20, 153-155.

This paper reported a research study conducted at 2 campuses, a State University College and a State University Center. Samples from both campuses completed an instrument which covered 3 broad areas: (1) characteristics of instructors who made the students "think"; (2) qualities of the instructors which students felt triggered their thought processes; and (3) the total number of instructors in contact with the 2 student samples, and how many of these made students think beyond the facts and data level. Some of the results were as follows. Students at both campuses often selected "thought provoking" instructors as being from outside their own academic major. The level of the course apparently had little to do with the stimulation of thought. The results showed exceptionally high correlation between the 2 sample groups in their choices of those qualities exhibited by instructors who made them think. About only one-fifth of the total number of teachers in contact with the samples were considered as "thought provokers" by their students. V. de Wolf. (D9, E4, F22, G4)

Walter, J. E. Relationships between selected values of students and their perception of a university instructor. *Dissertation Abstracts International*, 1972, *32*(8-A), 4387. (Ann Arbor, Michigan: University Microfilms. Order No. 72-1974.)

The purpose of the study was to investigate further the theoretical concept of relationships between personal variables of individuals and their perception of other persons. The study was an examination of relationships between values of university students, as measured by the Allport, Vernon, Lindzey Study of Values, and their perception of a university instructor, as measured on ten dimensions of behavior by the Purdue Rating Scale for Instruction. Other student variables were sex, grade expected for the course, major field of study, and seating location in the room. The findings of the study indicated that values of students did not appreciably affect their perception of instructor behavior, but that sex, grade expected, and major field of study were related to student perception of the behavior of the instructor. V. de Wolf. (C43, F6, F7, F11, F23, F75, K2)

Warrington, W. G. Student evaluation of instruction at Michigan State University. In A. L. Sockloff (Ed.), *Proceedings: The first invitational conference on faculty effectiveness as evaluated by students.* Philadelphia, Pa.: Measurement and Research Center, Temple University, 1973. Pp 164-182.

This report briefly described the development of the Student Instructional Rating System (SIRS) at Michigan State University. It also summarized the topics presented in four SIRS Research Reports. V. de Wolf. (A3, A4, A6, A9, B4, B28, C20, F20)

Weigel, R. G., Oetting, E. R., & Tasto, D. L. Differences in course grades and student ratings of teacher performance. *School and Society*, 1971, *99*, 60-62.

The present study was designed to determine whether a relationship between course grades and teacher evaluations was present for some instructors, but not for others, and also whether there were differences in this relationship for an instructor teaching different courses. Four instructors were rated by their students using a 23-item rating scale. Students were divided into 3 groups depending upon their final grade, A or B-C or D-F. Directions of the significant relationships between course grades and teacher evaluations varied between instructors. The most likely effect appeared to be a better evaluation of instructors by A students, with greater variability in evaluation by other grade groups, depending on the teacher and the class. The overall effects noted (A > B-C, A > D-F) suggested that the instructors in the present study geared their course toward the better students. Even though large correlational studies indicated that students' grades and evaluations of teachers were not importantly related, this relationship should not be dismissed lightly. V. de Wolf. (B80, F3, F5)

Whitely, S. E., Doyle, K. O., Jr., & Hopkinson, K. Student ratings and criteria for effective teaching. Report 731F. Minneapolis, Minnesota: Measurement Services Center, University of Minnesota, 1973.

Considerable concern has been generated in many quarters about the role of student ratings in the development of criteria for effective teaching. In the present study, a rating instrument was administered to students in a multi-section beginning language course. Both across- and between-sections data were studied for validity and generalizability. The results indicated that student ratings overlap with one of the principal criteria of teaching effectiveness, classroom achievement, but yet possess a unique component. The pattern of results further indicated that student ratings can contribute valuable information about teaching quality since the ratings are apparently generalizable and are based on criteria probably compatible with the criteria of those who must make decisions about instruction and instructors. Authors. (A6, B22, C60, F2, F6, F16, F19, F25, F27, F37, F62, F71, F76, G5, K2)

Whitlock, G. H., & Montgomery, J. R. Selection of outstanding teachers. Improving College and University Teaching, 1968, 16, 197-199.

The development of the Teacher Performance Checklist at the University of Tennessee, Knoxville Campus, and its use in the selection of outstanding teachers on this campus was documented. V. de Wolf. (A9, A10, B50, C37, K2)

Whitlock, L. G. The dimensions of observer perceptions of teacher performance. Dissertation Abstracts International, 1973, 33(8-B), 4006. (Ann Arbor, Michigan: University Microfilms. Order No. 73-2511.)

Students in a number of large classes completed a teacher behavior check list. An inverse factor analysis was performed on the student intercorrelation matrix for each of the five classes used in the study. This analysis yielded at least 10 factors per group. A cluster analysis was also performed. A number of clusters were found for each group and 2 of the clusters were found to be common across groups--cluster A corresponded to the second factor while cluster B corresponded to the first factor. In conclusion, an investigation was made of the relationship between cluster membership and 14 items reflecting learning preferences; no significant relationships were found. Similarly, no significant relationship was found between cluster membership and sex, college major, grade point average, and American College Test composite score. Finally, the relationship between cluster membership and students' evaluations of overall teacher performance was investigated and it was found that for 3 of the 5 teachers, a significant relationship existed. V. de Wolf. (A6, F5, F6, F7, F16)

Widlack, F. W., & Quereshi, M. Y. Student characteristics and instructor ratings: A person-perception approach. In C. Windle (Chair), Toward improved models of college teaching evaluation: Instructor, student, and organizational influences. Symposium presented at the American Psychological Association, Honolulu, September 1972.

In order to assess the validity of the Warr and Knapper information-processing model of person-perception, the students of a professor at Marquette University completed the Michill Adjective Rating Scale (MARS) at both the beginning and at the end of the course. In addition, the Rating Scale for Teachers (RST) was used at the final class session. To test the effects of the rater's sex and achievement level (AL) on the MARS and RTS ratings, a 2 x 2 multivariate analysis of variance design, in which sex and AL constituted the main effects, was employed. The rater's sex had a significant effect on all 4 stimulus persons' MARS scores. This fact suggested that there was a general person-perception phenomenon. On the other hand, the presence of a significant AL main effect for the MARS ratings of only the course instructor seemed to indicate, as expected, that the influence of the rater's AL was specifically directed at the perception of the teacher's personality. Finally, a high proportion of variance (42%) shared by the RST rating and the final MARS instructor rating implied that the perceived personality characteristics of the instructor were related to the way in which his/her students evaluated his/her teaching. V. de Wolf. (A6, A11, B68, C32, C33, D12, F3, F6, F12, F16, J5, J6)

Widlack, F. W., McDaniel, E. D., & Feldhusen, J. F. Factor analysis of an instructor rating scale. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, February 1973.

The purpose of the present study was threefold: (1) to determine the factor structure of a recently developed Course-Instructor Evaluation (CIE) instrument which had been widely used at Purdue University; (2) to ascertain the separability of the appraisal of the course from that of the instructor; and (3) to analyze the relationships among specific rating items and between specific items and global items, in order to assess the possible halo effect evoked by the CIE instrument. Three factors were found to underlie the CIE: Actor; Interactor; and Text-test. The overall course rating was found to be virtually inseparable from the overall instructor rating. The generally strong direct relationships among specific rating items and between specific items and global items implied a strong halo evoked by the CIE. This instrument was found to be a weak prospect for the effective diagnosis of courses and instructors. V. de Wolf. (A6, B8, C46, D28, G6)

Williams, V. G. Some contributions of student evaluation to course revision. Improving College and University Teaching, 1973, 21, 67-70.

Student reactions to an introductory counseling course at the University of Nebraska had been collected systematically by means of a questionnaire administered at the end of each semester. This article reported the results of a comparison between the evaluations obtained in 1968 and 1969. Several changes, stimulated for the most part by the outcome of the initial evaluation, were made in the form of the course between 1968 and 1969. A comparison of the 1968 and 1969 evaluations, along with a discussion of the implications of these comparisons, completed the report. V. de Wolf. (B46, D33, D34, J4)

Wilson, R. C., Dienst, E. R., & Watson, N. L. Characteristics of effective college teachers as perceived by their colleagues. *Journal of Educational Measurement*, 1973, *10*, 31-37.

The purpose of the research was to study the dimensions faculty members associate with being a good teacher. One hundred nineteen faculty members described the most effective and least effective teacher they knew using a set of 67 statements. A principal-components analysis of the descriptions of effective teachers produced 5 factors: *Research Activity and Recognition*, *Participation in the Academic Community*, *Intellectual Breadth*, *Relations with Students*, and *Concern for Teaching*. Only *Research Activity and Recognition* was significantly related to the academic rank and discipline of teachers nominated as effective. Rank, discipline, age and length of teaching experience of the nominators were unrelated to their descriptions of effective teachers. However, faculty with heavier teaching loads were more likely to assign high scores on *Concern for Teaching* and low scores on *Research Activity and Recognition* to the teacher they nominated as effective. Also, the effective teacher as seen by colleagues was contrasted with the effective teacher as seen by students. The two groups differed in the characteristics of effective teachers but agreed on who was effective and who was ineffective on this particular campus. V. de Wolf. (A6, B44, C38, D2, D3, D4, D10, H3)

Wofford, J. C. Attitudes and scholastic behavior. *The Journal of Educational Research*, 1968, *61*, 360-362.

The purpose of this study was to investigate the relationship between scholastic behavior and attitudinal variables and the interrelationship of scholastic behavior variables. Seventy-two students of general psychology answered a sentence completion attitude scale which measured attitudes toward two specific factors--instructor and the course, and two general factors--college and life. Scholastic behavior measures were absences, tardiness, and course grades. Attitudes toward life were identified as a suppressor and its variances statistically removed. Significant partial correlations were found between absences and both of the attitudes toward college and the composite of the four factors ($p < .05$). Significant relationships were also found for course grades with the attitudes toward the course and the composite scores. While tardiness scores were significantly negatively correlated with course grades, absence scores were not. Author. (A11, F3, F11, F21, F65)

Woolford, G. A. Teacher influence in a college of education. *Educational Research*, 1969, *11*, 148-152.

The reactions of students to different patterns of teacher influence were assessed in this study. For Spring term the two experimental courses (ECs) were student-centered. Five consecutive meetings of both ECs were tape recorded and analyzed. Immediately after the last meeting, the students completed a Post-Session Reaction Sheet (PSRS). During the Summer term a further 5 consecutive meetings of each group were analyzed. This time the course was instructor-centered. After the fifth meeting the PSRS was again completed by both ECs. Two control courses (CCs) completed the PSRS at the end of the school year. The reactions to limited participation of students in the ECs above and below the median on extroversion, stability, academic ability and sociometric status were compared. It was concluded that in general, students preferred patterns of teacher influence encouraging participation, regardless of their ability, sociometric status and personality traits. In comparing the ECs and the CCs on the PSRS both tended to be more satisfied with patterns of teacher influence encouraging participation. V. de Wolf. (C39, D16, F12, F13, F14, F16)

Yamamoto, K., & Dizney, H. F. College students' preferences among four types of professors. Journal of College Student Personnel, 1968, 9, 259-264.

Two instruments were administered to students: Eight College Professors, an 8-item, 5-choice, Likert-type rating scale constructed by the authors to investigate student preferences among 4 types of professors (Administrator, Socialite, Teacher, and Researcher); and the Allport, Vernon, and Lindzey Study of Values. The results of the present study, taken together with those of previous ones, seemed to suggest the generalizability of the overall pattern of students' preferences among the 4 professor types. At least among education students, Teacher received the highest ratings with a clear margin and Administrator the lowest regardless of students' sex. Researcher and Socialite fell somewhere in between and the exact order of these two depended upon the students' class level and major. While the general pattern of professor image was replicated with certainty, the authors' efforts to find some personal value correlates of the preferential ratings failed. V. de Wolf. (A11, B18, D40, F2, F6, F7, F11, F66)

Yamamoto, K., Smallings, M. S., & Wiersma, J. Students perceptions of Eight Professors in small colleges. The Journal of Experimental Education, 1972, 41(1), 91-96.

Responses on a Likert-type instrument, the Eight College Professors (Form B), from 664 education juniors and seniors in five small colleges were analyzed to examine the preferential rankings given four types of social science professors as an instructor. While some suggestive variations were observed, the overall order of preference was Teacher--Socialite--Researcher--Administrator (the least preferred). This was in contrast to earlier findings on university students who placed Researcher ahead of socialite. Female professors were consistently less preferred to male professors by college students of both sexes. Authors. (A11, C40, D1, F6, F22, F66)

Yonge, G. D., & Sassenrath, J. M. Student personality correlates of teacher ratings. Journal of Educational Psychology, 1968, 59, 44-52.

The 227 students taught by 3 instructors of a course in educational psychology completed an instructor/course rating form and personality test. The ratings were factored and 9 factors were rotated by the normalized varimax technique. Factor scores were computed for each student. Correlations between the personality test plus course grades and factor scores were computed separately for each of the 3 instructor samples. Differences between correlations were tested for significance among the 3 instructors. The results indicate that generally different personality scores correlate with a given factor score, and in some instances the same personality variable correlates with a given factor score in opposite directions from one instructor to another. It was concluded that the factors have a different psychological meaning for each instructor. The results were discussed in terms of the known differences among the instructors. V. de Wolf. (A6, F3, F12)

Yongkittkul, C., Gillmore, G. M., & Brandenburg, D. C. Does the time of course meeting affect course ratings by students? Research Report #346. Urbana, Illinois: Measurement and Research Division, Office of Instructional Resources, University of Illinois, 1974.

The ADVISOR (revised form) was used to obtain student ratings of instructors and courses for classes offered in spring 1971. The classes were divided according to nine groups according to the beginning time of class meeting in order to assess possible differential ratings by time.

The first two subscales of the ADVISOR, Overall Evaluation of Course and Instructor, yielded significant overall F-ratios from a one-way analysis of variance, but none of the individual comparisons among the nine class meeting times were significant. The other two subscales did not yield significant overall F-ratios. By utilizing the above results and ω^2 to assess the degree of relationship between ratings and time of class meeting, it was concluded that the time of day at which classes meet does not appreciably affect instructor or course ratings. Authors. (B3, C6, G2)

Zelby, L. W. Student-faculty evaluation. Science, 1973, 183, 1267-1270.

In this article the author attempted to demonstrate the importance of format of a student ratings evaluation questionnaire by showing that, given a specific format, it was possible to adapt one's teaching technique to obtain a good or bad evaluation, and that a good evaluation may be associated with a teaching technique of lesser educational value than a poor evaluation. Essentially what the author did was to teach two sections of two courses, keeping the administrative procedures for each course constant but differing the teaching technique in the sections of each course. He then discussed the results of the student evaluations of the courses--it seemed that students gave higher ratings to the "more traditionally taught course" and lower ratings to the "more experimentally taught course." He concluded that good student evaluation ratings may not be consistent with the best educational practices and that the particular questionnaire used at his institution was conducive to stereotyped teaching. He saw the potential dangers of student ratings of faculty in decreasing the effectiveness of education lying not in the evaluation process, but in the format of the evaluation. V. de Wolf. (D16)

APPENDICES

A. Research related to the development, construction, and validation of the student ratings questionnaires themselves

1. Survey of various institutions with respect to their use(s) of student ratings of instruction
2. Review of student ratings of instruction research across the country; bibliography of student ratings articles
3. Comparison(s) of different types of scaling of rating forms (ex., graphic vs. forced-choice or bipolar vs. unipolar)
4. Use of a variety of rating questionnaires or of systems of weighting items uniquely for different departments within a university
5. The calibre of student ratings when used for different purposes (ex., the differences between scores on rating questionnaires when for the personal use of the professor and when used strictly for research purposes)
6. Factor analysis of rating questionnaires
7. Bias in student ratings of instruction
8. Discriminant analysis of student ratings questionnaire; discriminant function loadings
9. Actual development and construction of instruments for student evaluation of instruction
10. Determining the characteristics differentiating "best" from "worst" college instructors or "effective" from "ineffective" professors
11. Reliability of ratings questionnaires; generalizability theory and student ratings
12. Inter-university comparisons of a particular student ratings questionnaire
13. Norming; approximating decile norms
14. Face validity
15. Differential weighting of items on one particular student ratings questionnaire
16. Cross-cultural scaling research
17. Internal consistency

B. Institution rated in the research study

1. Univ. of Washington
2. Western Washington State College
3. University of Illinois
4. Michigan State Univ.
5. Univ. of Calif., Berkeley
6. Old Dominion Univ.
7. Univ. of Northern Colorado (Formerly Colorado St. Coll.)
8. Purdue Univ.
9. Indiana Univ.
10. Penn. State Univ.
11. Bowling Green St. Univ.
12. Univ. of North Carolina
13. Univ. of Arizona, Tuscon
14. Western Michigan Univ.
15. Univ. of Delaware
16. Tufts Univ.
17. Univ. of Wisconsin, Milwaukee
18. Univ. of Iowa
19. Univ. of Michigan
20. Univ. of Miami
21. Univ. of Texas
22. Univ. of Minnesota, Morris; Univ. of Minn., (unspecified)
23. Southern Illinois Univ.
24. Univ. of New England, Armidale, Australia
25. Cleveland State Univ.
26. Northwestern Univ.
27. Sheridan College Sheridan, Wyoming
28. Florida State Univ.
29. Univ. of New Mexico, Las Cruces
30. Western Kentucky Univ.
31. Northern Illinois Univ.
32. Iowa State Univ.
33. Univ. of South Florida
34. State Univ. College of N. Y., New Paltz
35. Ohio State Univ.
36. Univ. of Utah
37. Oregon State Univ.
38. Yale Univ.
39. Univ. of Wisconsin, Green Bay
40. Texas A & M Univ.
41. Carnegie-Mellon Univ.
42. Austin College
43. San Jose State College
44. Univ. of Calif., Davis
45. Kansas State Univ.
46. Univ. of Nebraska, Omaha; Univ. of Nebraska (unspecified)
47. Northampton County Area Community College
48. Illinois State Univ.
49. Univ. of Cincinnati
50. Univ. of Tennessee, Knoxville; Univ. of Tennessee (unspecified)
51. Centre College of Kentucky
52. Calif. State College, Long Beach
53. Brooklyn College
54. Grinnell College
55. Harcum Junior College, Bryn Mawr, Pa.
56. Antioch
57. Univ. of Colorado
58. Oberlin
59. Miami University, Ohio
60. Columbia Univ.
61. Univ. of Hawaii
62. Baldwin-Wallace College, Ohio
63. Univ. of Windsor, Canada
64. Univ. of Western Ontario, Canada
65. Kent State Univ.
66. Univ. of Toledo, Toledo, Ohio
67. Wayne State Univ.
68. Marquette Univ.
69. Brigham Young Univ.
70. Temple Univ. Pa.
71. Univ. of Maryland
72. Univ. of Bradford (England)
73. Maritime Academy
74. Shippensburg State Univ.
75. Eureka College
76. Lake Land College
77. Univ. of Alabama
78. Univ. of S. Alabama
79. Freed Hardeman College
80. Colorado State Univ.
81. Broome Community College N.Y.
82. Ball State Univ.
83. State College of Iowa
84. Murray State Univ., Murray, Kentucky
85. Lee College

C. Specific rating forms employed or modifications of specific rating forms

Abbreviation	Name of form and/or author(s)/user(s)
IU	1. Indiana University's Student Evaluation of Instruction
FE	2. Faculty Evaluation Survey
SIR	3. ETS's Student Instructional Report (SIR)
CIS	4. Course-Instructor Survey (CIS) General questionnaire of Albino and Liberty
CEQ	5. Course Evaluation Questionnaire (CEQ) at University of Illinois
----	6. The Advisor at University of Illinois
SIRR	7. Student Instructional Rating Report (SIRR) at Univ. of Michigan
----	8. Forms based upon the work of Isaacson et al. (1963, 1964)*(such as: Rees' modification called the College Instructor Evaluation Inventory [CIEI]; or the McKeachie-Lin Student Opinion Questionnaire [SOQ])
CEQ	9. Student Course Evaluation Questionnaire (CEQ) at Univ. of Wisconsin
----	10. Reading Validity Index used at Univ. of Wisconsin
TRS	11. Teacher Rating Scale (TRS) of Kilpatrick and Cantril used by Dedrick
TDI	12. Teacher Description Instrument (TDI) of Despande, Webb, and Marks
IIQ	13. Instructional Improvement Questionnaire (IIQ) at Southern Illinois Univ.
----	14. Course and Instructor Questionnaire at Cleveland State Univ.
----	15. Students' View of an Ideal Professor of Gadzella
----	16. Instructor Evaluation Scale
SECTB	17. Student Evaluation of College Teaching Behavior (SECTB)
----	18. Student Survey Instrument
TAB	19. Teaching Assessment Blank (TAB) of Holmes at Univ. of Texas
SIRS	20. Student Instructional Rating System (SIRS) at Michigan State Univ.
IEQ	21. Instructor Evaluation Questionnaire (IEQ) at Univ. of Nebraska, Omaha
IRQ	22. Instructor Rating Questionnaire (IRQ) at Northampton County Area Community College
----	23. Teacher Performance Checklist of Douglass used by Lucas
----	24. Osgood's Semantic Differential and derivatives (such as Rees' College Instructor Rating Scale [CIRS])
----	25. Cognitive Structure Questionnaire
----	26. A Student's Rating Scale of an Instructor of Eidsmoe used by Meredith
SSOT	27. Survey of Student Opinion of Teaching (SSOT) of Miller
----	28. Student Perception of Teaching Effectiveness Scale of Perry and Baumann
----	29. Plant, Miller, and Sawrey Scale used by Plant and Sawrey
----	30. Pfeiffer and Rosbach's 30-item teaching activity stimuli (Pfeiffer at LaSalle College)
----	31. Student Report Form of Caffrey
RST	32. Rating Scale for Teachers (RST) of Quereshi at Marquette Univ.
MARS	33. Mitchell Adjective Rating Scale (MARS) used by Quereshi
----	34. Student Scale of Rayder
----	35. Course Evaluation Survey of Cohen used by Rosenshine, Cohen, and Furst

C. Specific rating forms employed or modifications of specific rating forms (continued)

Abbreviation	Name of form and/or author(s)/user(s)
----	36. Indiana University's Student Evaluation of Faculty of Goldenbaum and Wheeler used by Stallings and Singhal
----	37. Teacher Performance Checklist of Whitlock and Montgomery
----	38. Colleague Teacher-Description Scales of Wilson, Dienst, and Watson
----	39. Post-Session Reaction Sheet of Woolford
----	40. Eight College Professors of Yamamoto
----	41. Analysis of College Teaching of Bridges, Brown, Greenwood, and Ware
----	42. Student Evaluation Form, Part I, at Univ. of Minnesota
PRSI	43. Purdue Rating Scale for Instruction (PRSI) and modifications (such as Bausell's or Price's)
PIPI	44. Purdue Instructor Performance Indicator (PIPI)
IBDQ	45. Purdue's Instructor Behavior Description Questionnaire (IBDQ)
CIE	46. Purdue's Course-Instructor Evaluation (CIE) form
----	47. Purdue's Cafeteria System
TPI	48. Teacher Performance Inventory (TPI) at Iowa State Univ.
----	49. Rating Scale at Univ. of Northern Colorado
----	50. Faculty Evaluation Checklist at Harcum Junior College
----	51. Indiana University's College of Arts & Sciences course evaluation instrument
----	52. Indiana University's School of Business course evaluation instrument
----	53. Mathematical Sciences Instructor Evaluation Form at Ball State Univ.
----	54. Registration Scale at Univ. of Washington
----	55. Univ. of Illinois, Urbana-Champaign Physics Form
----	56. Teacher Description Scales of Hildebrand & Wilson (38-item)
ISPI	57. Inventory of Student Perceptions of Instruction (ISPI) used by Scott, Halpin & Schnittjer
----	58. Forms modified from Fiedler's (1967)** adjective rating scales for leadership effectiveness such as, Turner's Instructor Preference Scale
CAQ	59. Course Attitude Questionnaire (CAQ) by Spencer (this is the bases for the Univ. of Illinois CEQ, see #C5)
SOS	60. Student Opinion Survey at Univ. of Minnesota

*Isaacson, R. L., McKeachie, W. J., & Milholland, J. E. Correlates of teacher personality variables and student ratings. *Journal of Educational Psychology*, 1963, *54*, 110-117.

Isaacson, R. L., McKeachie, W. J., Milholland, J. E., Lin, Y. G., Hofeller, M., Baerwaldt, J. W., & Zinn, K. L. Dimensions of student evaluations of teaching. *Journal of Educational Psychology*, 1964, *55*, 344-351.

**Fiedler, F. *A theory of leadership effectiveness*. New York: McGraw Hill, 1967.

D. Student ratings of instruction based upon grouping faculty by:

1. Sex
2. Number of years of teaching experience
3. Age
4. Rank
5. Those with evening school teaching experience vs. those without it
6. General research productivity; amount of research as measured by number of publications (including books, books edited, articles, technical reports, book reviews)
7. Research measure--number of awards received (both government and institution awards)
8. Tenure vs. non-tenure
9. Highest degree attained; years of formal education
10. General teaching area or division within the institution; major area
11. Research measure--authorship position (1st, 2nd, etc., author of books, articles, reports, booklets, funded grants)
12. Personality (ex. conceptual system orientation, empathy, perceptual organization); perception of teacher personality
13. Membership in prestigious professional organizations
14. Severity of grading practice ("grading leniency")
15. Allocation of time distributed among academic activities; time expenditure among various classroom activities
16. Teaching methodology; teaching style; instructional model underlying course
17. Stress measures
18. Overall knowledge of subject being taught; competence
19. Overall knowledge of correct teaching practices or perception of the teaching task
20. Ability to use good methods of oral communication
21. Salary
22. Whether or not the professor is teaching in terms of objectives outlined at the beginning of the course
23. Whether professor has full-time or part-time status
24. Perception of level of learning
25. Openness to change; openness to experience
26. Perception of others and self
27. Number of years at the institution
28. Reputation (i.e., teacher folklore; "halo" effect)
29. Credibility
30. Attitudes towards student ratings of instruction
31. Research measure--number of papers presented at meetings of scientific/professional societies
32. Research measure--number of BA honors, MA, or MS thesis or Ph.D. dissertation committees chaired
33. Availability of instructor to students in his/her class
34. Explicitness of course policies/assignments
35. Appearance
36. Graduate grade point average
37. Verbal self disclosure
38. Tactual body accessibility

E. Student ratings and the similarities between professor and/or course and student and/or alumni/ae in the following area:

1. Cognitive compatibility; cognitive style
2. Attitude similarity
3. Major field, i.e., is professor a member of faculty in student's major department?
4. Major field, i.e., is course in student's major field?
5. Areas of agreement on qualities of "best teacher," "best student"
6. Belief congruence
7. Agreement on teacher appearance behavior trait

F. Student ratings of instruction based upon grouping students by:

1. Age ("rater maturity")
2. Class level; year in college; level of education (ex. undergraduate vs. graduate or freshman-sophomore vs. junior-senior); term
3. Course grade(s) (course grades could = "grade up to now" + final grade)
4. Type of grade(s) previously received by students from the same professor
5. Cumulative gpa
6. Sex
7. Major; division within the university; curriculum
8. Veteran vs. non-veteran
9. Number of courses previously taken by students in this same department
10. How well course fitted into schedule
11. Personal values of students (ex. attitudes towards instruction in general; college goals, in general, or objectives; values as measured by ALV Study of Values)
12. Personality traits (ex. as measured by EPPS, anxiety scales such as Taylor's , Alpert-Haber's or Sarason's, Eysenck Personality Inventory, Rokeach's dogmatism scale, Omnibus Personality Inventory, creativity self-rating scores, or memory measures)
13. Sociometric or socioeconomic status
14. Perception of classroom atmosphere; "feeling of control" in classroom
15. Plans with regards to graduate school (ex. undergraduates planning to go to graduate school vs. those who are not; those going vs. those uncertain vs. those not going)
16. Scholastic aptitude, ability, or achievement (i.e., relationship of low-ability students, high-ability students and instructors/courses most effective in stimulating each group; high-achievers vs. low-achievers; over-achievers vs. under-achievers; scores achieved on standard, non-course related aptitude tests such as SAT, Miller Analogies, ACT, Otis IQ, Nelson-Denny Reading tests; course pretests)
17. Was course recommended by another student?
18. Comparison of ratings made by regularly admitted students with those of "specially admitted students" (i.e., culturally different or culturally disadvantaged)
19. Perception of level of learning or of understanding of subject matter; getting as much as anticipated from the course?
20. Anonymous vs. identified student responses
21. Number of class absences; class attendance
22. Inter-institutional comparisons

F. Student ratings of instruction based upon grouping students by: (cont.)

23. Expected course grade
24. Difference between expected and actual course grades
25. Results on a standardized or common final exam (i.e., any uniform measure of amount learned)
26. Expectation of difficulty of course; amount of work demanded of students by course
27. Amount of interest in course material; liking for subject matter of course
28. Number of credits taken this term or quarter
29. Preconception of enjoyment which would be received from course
30. Transfer vs. non-transfer students
31. Completion vs. incompletion of student teaching experience
32. Expected contribution of course to planned vocation
33. Expected contribution of course to general education
34. Expected ease of course
35. Amount of independent work done by student
36. Grade deserved
37. Effort expended during course, how hard student tried to learn
38. Importance of the course
39. Most important objective(s) of teaching as favored by students; the value of certain teaching performance factors
40. Extracurricular participation
41. Preference for learning environment (ex. independent study vs. lecture; teacher-centered vs. student-centered)
42. Difference between predicted course grade and actual grade
43. Those who prepared assignments daily vs. those who didn't
44. Those who prepared for class vs. those who didn't
45. Those who voluntarily participated in class vs. those who didn't
46. Those who sought out explanations for materials not understood vs. those who didn't
47. Those who looked up minor points not understood in class vs. those who didn't
48. Those who monopolized class discussions vs. those who didn't
49. Those who made relationships between new and old material and those who didn't
50. Those who made relationships between this course and other courses and those who didn't
51. Home background environment (ex. urban or non-urban)
52. Marital status
53. Score on the Minnesota Teacher Attitude Inventory
54. Seriousness of academic intent while not in class
55. Number of pages of reading read
56. Number of hours of study per week
57. Reason for course selection
58. Home stability (i.e., a roundabout assessment of marital status of parents--"stable" vs. "broken," that is, parents are divorced)
59. Number of years residing in the state in which the college is located
60. Amount of travel experience (U. S., foreign)
61. Race
62. Attitudes towards the value of student ratings
63. Measurement on a vocational interest inventory (ex. California Interest Inventory)

F. Student ratings of instruction based upon grouping students by: (cont.)

64. Generalized course and instructor ratings taken at the beginning of course
65. Tardiness
66. Preference for and perception of different professor "types"
67. Income, family
68. High school grade point average
69. Goals--educational
70. Goals--vocational
71. Liking for professor as a person
72. Life history factors
73. Does course fulfill student needs?
74. Does course fulfill expectations? (i.e., general expectations)
75. Seating location in classroom
76. Conscientiousness of students in filling out student ratings questionnaire

G. Student ratings of instruction based upon certain characteristics of the course itself, such as:

1. Class size (ex. comparing ratings of different sections of same course, each section having different class size; dividing all returned student ratings forms into one of three groups, depending on whether the class size is small, medium, or large)
2. Time of day the course is taught
3. Field of the course
4. Course level
5. Required vs. elective
6. Course ratings compared with faculty ratings
7. Instructional methodology
8. Assessment of students' progress on relevant course objectives compared with the instructors' ratings of progress in these objectives.

H. Comparison of student ratings of instruction and other types of evaluative ratings such as:

1. Faculty self-ratings; self-evaluation of teaching effectiveness by faculty; self-evaluation of overall contribution to college
2. Administrators' ratings of faculty
3. Colleague ratings of both real and ideal faculty characteristics; peer group rating of teaching effectiveness of professors or of overall contribution to college
4. Ratings of videotape excerpts of actual classroom teaching of faculty--ratings are done by expert judges
5. Alumni/ae ratings
6. Trained classroom observers--general
7. The ratings of expert observers who used Flanders Interaction Analysis Categories to assess professor's teaching behavior
8. Supervisor's ratings (i.e., rating of chairman/woman)

I. Ideal ratings of instruction by students and relationship(s) of real and ideal ratings

1. Comparisons of student ratings of faculty/course with their ratings of the ideal faculty member or ideal course conditions
2. Comparison of the ideal professor in general and the ideal professor in a variety of fields
3. Views of the ideal professor as seen by various students grouped by personality type
4. Characteristics of the ideal professor and/or the ideal instructional situation

J. Documentation of changes provoked by use of student ratings

1. The effect of publishing student ratings upon subsequent ratings of the course and/or instructor
2. Changes in instructor self-ratings resulting from feedback of student ratings
3. Changes in instruction (noticed after instructor received feedback from student ratings; noticed in subsequent terms on student ratings after individual conference and discussion of earlier student ratings)
4. Contribution(s) of student ratings to course revision(s)
5. Test-retest use of student ratings (ex. a comparison of student ratings of an instructor while they had the professor and two years later after completing subsequent coursework--the students were still in same college; a comparison of student ratings with ratings of alums taken 5 or 10 years after graduating; relationship of relative importance attached to the various rating traits by students and alums; beginning and end of quarter comparisons)
6. Changes in student attitudes or perceptions during the course of instruction
7. Correlation of student ratings and the probability of taking another course in the same department

K. Miscellaneous topics concerned with student ratings of instruction

1. American Council on Education (ACE ratings of graduate departments correlated with student ratings
2. Prediction of student ratings
3. Student ratings and the selection of "outstanding faculty members" (ex. the relationship between items differentiating instructors rated highly by students and instructors receiving low ratings from students with those items differentiating the top group of faculty nominated for an outstanding teaching award from another group of faculty randomly selected from among those who were not nominated for the outstanding teaching award)
4. Comparison of student ratings of two groups of graduate teaching assistants--those who participated in a teaching seminar and those who did not
5. Student ratings of the images of 3 academic occupations vs. 2 nonacademic occupations